```
=> fil capl;d que l12; d que l13
FILE 'CAPLUS' ENTERED AT 12:03:32 ON 15 AUG 2003
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
COPYRIGHT (C) 2003 AMERICAN CHEMICAL SOCIETY (ACS)

Copyright of the articles to which records in this database refer is
held by the publishers listed in the PUBLISHER (PB) field (available
for records published or updated in Chemical Abstracts after December
26, 1996), unless otherwise indicated in the original publications.
The CA Lexicon is the copyrighted intellectual property of the
American Chemical Society and is provided to assist you in searching
databases on STN.  Any dissemination, distribution, copying, or storing
of this information, without the prior written consent of CAS, is
strictly prohibited.

FILE COVERS 1907 - 15 Aug 2003  VOL 139 ISS 8
FILE LAST UPDATED: 14 Aug 2003  (20030814/ED)

  This file contains CAS Registry Numbers for easy and accurate
  substance identification.
```

inventors search

```
L1          102 SEA FILE=CAPLUS ABB=ON  BLANKENBECLER R?/AU
L2           80 SEA FILE=CAPLUS ABB=ON  OHLSSON M?/AU
L3         1196 SEA FILE=CAPLUS ABB=ON  PETERSON C?/AU
L4           22 SEA FILE=CAPLUS ABB=ON  RINGNER M?/AU
L7        34837 SEA FILE=CAPLUS ABB=ON  ALGORITHM/CT
L9       739531 SEA FILE=CAPLUS ABB=ON  PROTEINS/CW
L10       22499 SEA FILE=CAPLUS ABB=ON  L9(L)(STRUCTURE? OR ALIGN?)
L12           2 SEA FILE=CAPLUS ABB=ON  (L1 OR L2 OR L3 OR L4) AND L10 AND L7


L1          102 SEA FILE=CAPLUS ABB=ON  BLANKENBECLER R?/AU
L2           80 SEA FILE=CAPLUS ABB=ON  OHLSSON M?/AU
L3         1196 SEA FILE=CAPLUS ABB=ON  PETERSON C?/AU
L4           22 SEA FILE=CAPLUS ABB=ON  RINGNER M?/AU
L9       739531 SEA FILE=CAPLUS ABB=ON  PROTEINS/CW
L10       22499 SEA FILE=CAPLUS ABB=ON  L9(L)(STRUCTURE? OR ALIGN?)
L11           7 SEA FILE=CAPLUS ABB=ON  (L1 OR L2 OR L3 OR L4) AND L10
L13           1 SEA FILE=CAPLUS ABB=ON  L11 AND NONRAND?/TI

=> s l12 or l13

L155          3 L12 OR L13

=> fil wpids; d que l36; d que l43

FILE 'WPIDS' ENTERED AT 12:03:33 ON 15 AUG 2003
COPYRIGHT (C) 2003 THOMSON DERWENT

FILE LAST UPDATED:            13 AUG 2003      <20030813/UP>
MOST RECENT DERWENT UPDATE:      200352         <200352/DW>
DERWENT WORLD PATENTS INDEX SUBSCRIBER FILE, COVERS 1963 TO DATE

>>> NEW WEEKLY SDI FREQUENCY AVAILABLE --> see NEWS  <<<

>>> PATENT IMAGES AVAILABLE FOR PRINT AND DISPLAY <<<
```

```
>>> FOR DETAILS OF THE PATENTS COVERED IN CURRENT UPDATES,
    SEE http://www.derwent.com/dwpi/updates/dwpicov/index.html <<<

>>> FOR A COPY OF THE DERWENT WORLD PATENTS INDEX STN USER GUIDE,
    PLEASE VISIT:
 http://www.stn-international.de/training_center/patents/stn_guide.pdf <<<

>>> FOR INFORMATION ON ALL DERWENT WORLD PATENTS INDEX USER
    GUIDES, PLEASE VISIT:
    http://www.derwent.com/userguides/dwpi_guide.html <<<


L32           11 SEA FILE=WPIDS ABB=ON  BLANKENBECLER R?/AU
L33           11 SEA FILE=WPIDS ABB=ON  OHLSSON M?/AU
L34          225 SEA FILE=WPIDS ABB=ON  PETERSON C?/AU
L35            1 SEA FILE=WPIDS ABB=ON  RINGNER M?/AU
L36            1 SEA FILE=WPIDS ABB=ON  L35 AND (L32 OR L33 OR L34)


L32           11 SEA FILE=WPIDS ABB=ON  BLANKENBECLER R?/AU
L33           11 SEA FILE=WPIDS ABB=ON  OHLSSON M?/AU
L34          225 SEA FILE=WPIDS ABB=ON  PETERSON C?/AU
L35            1 SEA FILE=WPIDS ABB=ON  RINGNER M?/AU
L39       114374 SEA FILE=WPIDS ABB=ON  PROTEIN#
L40         2132 SEA FILE=WPIDS ABB=ON  L39(5A)(STRUCTURE# OR CONFORM? OR
                ALIGN?)
L43            1 SEA FILE=WPIDS ABB=ON  (L32 OR L33 OR L34 OR L35) AND L40


=> s l36 or l43

L156          1 L36 OR L43

=> fil medl; d que l53; d que l66

FILE 'MEDLINE' ENTERED AT 12:03:35 ON 15 AUG 2003

 FILE LAST UPDATED: 14 AUG 2003 (20030814/UP).  FILE COVERS 1958 TO DATE.

 On April 13, 2003, MEDLINE was reloaded.  See HELP RLOAD for details.

 MEDLINE thesauri in the /CN, /CT, and /MN fields incorporate the
 MeSH 2003 vocabulary.  See http://www.nlm.nih.gov/mesh/changes2003.html
 for a description on changes.

 This file contains CAS Registry Numbers for easy and accurate
 substance identification.


L53            0 SEA FILE=MEDLINE ABB=ON  BLANKENBECLER R?/AU


L54           39 SEA FILE=MEDLINE ABB=ON  OHLSSON M?/AU
L55         1382 SEA FILE=MEDLINE ABB=ON  PETERSON C?/AU
L56           17 SEA FILE=MEDLINE ABB=ON  RINGNER M?/AU
L58       130531 SEA FILE=MEDLINE ABB=ON  PROTEIN CONFORMATION+NT/CT
L64        36956 SEA FILE=MEDLINE ABB=ON  ALGORITHMS/CT
L65        48459 SEA FILE=MEDLINE ABB=ON  SOFTWARE+NT/CT
L66            0 SEA FILE=MEDLINE ABB=ON  (L54 OR L55 OR L56) AND L58 AND (L64
                OR L65)
```

```
=> fil embase; d que l82; d que l90; d que l93
```

FILE 'EMBASE' ENTERED AT 12:03:36 ON 15 AUG 2003

FILE COVERS 1974 TO 14 Aug 2003 (20030814/ED)

EMBASE has been reloaded. Enter HELP RLOAD for details.

This file contains CAS Registry Numbers for easy and accurate substance identification.

```
L82           0 SEA FILE=EMBASE ABB=ON  BLANKENBECLER R?/AU


L81      219472 SEA FILE=EMBASE ABB=ON  PROTEIN STRUCTURE+NT/CT
L83          38 SEA FILE=EMBASE ABB=ON  OHLSSON M?/AU
L84        1039 SEA FILE=EMBASE ABB=ON  PETERSON C?/AU
L85          16 SEA FILE=EMBASE ABB=ON  RINGNER M?/AU
L86         303 SEA FILE=EMBASE ABB=ON  ATOM?(3A)DISTANCE#
L87           2 SEA FILE=EMBASE ABB=ON  BINARY ASSIGNMENT#
L88         276 SEA FILE=EMBASE ABB=ON  MEAN FIELD#
L89         487 SEA FILE=EMBASE ABB=ON  ENERGY FUNCTION#
L90           1 SEA FILE=EMBASE ABB=ON  L81 AND (L83 OR L84 OR L85) AND (L86
                OR L87 OR L88 OR L89)


L81      219472 SEA FILE=EMBASE ABB=ON  PROTEIN STRUCTURE+NT/CT
L83          38 SEA FILE=EMBASE ABB=ON  OHLSSON M?/AU
L84        1039 SEA FILE=EMBASE ABB=ON  PETERSON C?/AU
L85          16 SEA FILE=EMBASE ABB=ON  RINGNER M?/AU
L91       22056 SEA FILE=EMBASE ABB=ON  ALGORITHM/CT
L92       23728 SEA FILE=EMBASE ABB=ON  COMPUTER PROGRAM/CT
L93           1 SEA FILE=EMBASE ABB=ON  L81 AND (L83 OR L84 OR L85) AND (L91
                OR L92)

=> s l90 or l93

L157          1 L90 OR L93

=> fil PASCAL, BIOTECHNO, ESBIOBASE, LIFESCI, BIOSIS, TOXCENTER, scisearch
```

FILE 'PASCAL' ENTERED AT 12:03:37 ON 15 AUG 2003

FILE 'BIOTECHNO' ENTERED AT 12:03:37 ON 15 AUG 2003

FILE 'ESBIOBASE' ENTERED AT 12:03:37 ON 15 AUG 2003

FILE 'LIFESCI' ENTERED AT 12:03:37 ON 15 AUG 2003

```
FILE 'BIOSIS' ENTERED AT 12:03:37 ON 15 AUG 2003
COPYRIGHT (C) 2003 BIOLOGICAL ABSTRACTS INC.(R)

FILE 'TOXCENTER' ENTERED AT 12:03:37 ON 15 AUG 2003
COPYRIGHT (C) 2003 ACS

FILE 'SCISEARCH' ENTERED AT 12:03:37 ON 15 AUG 2003
COPYRIGHT 2003 THOMSON ISI

=> d que l118

L1          102 SEA FILE=CAPLUS ABB=ON  BLANKENBECLER R?/AU
L2           80 SEA FILE=CAPLUS ABB=ON  OHLSSON M?/AU
L3         1196 SEA FILE=CAPLUS ABB=ON  PETERSON C?/AU
L4           22 SEA FILE=CAPLUS ABB=ON  RINGNER M?/AU
L106         93 SEA L1
L107        195 SEA L2
L108       7425 SEA L3
L109         80 SEA L4
L110     355959 SEA PROTEIN#(5A)(STRUCTUR? OR ALIGN? OR CONFORM?)
L111       5034 SEA ATOM?(3A) DISTANCE#
L118          3 SEA (L106 OR L107 OR L108 OR L109) AND L110 AND ((L111 OR L112
                OR L113 OR L114) OR L117) AND (L115 OR L116)


=> fil uspatf; d que l142; d que l143

FILE 'USPATFULL' ENTERED AT 12:03:40 ON 15 AUG 2003
CA INDEXING COPYRIGHT (C) 2003 AMERICAN CHEMICAL SOCIETY (ACS)

FILE COVERS 1971 TO PATENT PUBLICATION DATE: 14 Aug 2003 (20030814/PD)
FILE LAST UPDATED: 14 Aug 2003 (20030814/ED)
HIGHEST GRANTED PATENT NUMBER: US6606748
HIGHEST APPLICATION PUBLICATION NUMBER: US2003154532
CA INDEXING IS CURRENT THROUGH 14 Aug 2003 (20030814/UPCA)
ISSUE CLASS FIELDS (/INCL) CURRENT THROUGH: 14 Aug 2003 (20030814/PD)
REVISED CLASS FIELDS (/NCL) LAST RELOADED: Jun 2003
USPTO MANUAL OF CLASSIFICATIONS THESAURUS ISSUE DATE: Jun 2003

>>>  USPAT2 is now available.  USPATFULL contains full text of the     <<<
>>>  original, i.e., the earliest published granted patents or        <<<
>>>  applications.  USPAT2 contains full text of the latest US         <<<
>>>  publications, starting in 2001, for the inventions covered in     <<<
>>>  USPATFULL.  A USPATFULL record contains not only the original     <<<
>>>  published document but also a list of any subsequent             <<<
>>>  publications.  The publication number, patent kind code, and     <<<
>>>  publication date for all the US publications for an invention    <<<
>>>  are displayed in the PI (Patent Information) field of USPATFULL   <<<
>>>  records and may be searched in standard search fields, e.g., /PN, <<<
>>>  /PK, etc.                                                         <<<

>>>  USPATFULL and USPAT2 can be accessed and searched together        <<<
>>>  through the new cluster USPATALL.  Type FILE USPATALL to          <<<
>>>  enter this cluster.                                               <<<
>>>                                                                    <<<
>>>  Use USPATALL when searching terms such as patent assignees,       <<<
>>>  classifications, or claims, that may potentially change from     <<<
>>>  the earliest to the latest publication.                           <<<

This file contains CAS Registry Numbers for easy and accurate
substance identification.
```

```
L131          11 SEA FILE=USPATFULL ABB=ON  BLANKENBECLER R?/AU
L132           5 SEA FILE=USPATFULL ABB=ON  OHLSSON M?/AU
L133         307 SEA FILE=USPATFULL ABB=ON  PETERSON C?/AU
L134           1 SEA FILE=USPATFULL ABB=ON  RINGNER M?/AU
L135        1604 SEA FILE=USPATFULL ABB=ON  PROTEIN#(5A)(STRUCTUR? OR ALIGN? OR
                 CONFORM?)/IT,TI,AB,CLM
L136        2262 SEA FILE=USPATFULL ABB=ON  ATOM?(3A)DISTANCE# OR (ATOM?(3A)DIST
                 ANCE#)/IT
L137          32 SEA FILE=USPATFULL ABB=ON  BINARY ASSIGNMENT# OR (BINARY
                 ASSIGNMENT#)/IT
L138         228 SEA FILE=USPATFULL ABB=ON  MEAN FIELD# OR (MEAN FIELD#)/IT
L139         954 SEA FILE=USPATFULL ABB=ON  ENERGY FUNCTION? OR (ENERGY
                 FUNCTION?)/IT
L140      159150 SEA FILE=USPATFULL ABB=ON  ALGORITH? OR ALGORITH?/IT
L141      615026 SEA FILE=USPATFULL ABB=ON  COMPUT? OR COMPUT?/IT
L142           1 SEA FILE=USPATFULL ABB=ON  (L131 OR L132 OR L133 OR L134) AND
                 L135 AND (L136 OR L137 OR L138 OR L139 OR L140 OR L141)


L131          11 SEA FILE=USPATFULL ABB=ON  BLANKENBECLER R?/AU
L132           5 SEA FILE=USPATFULL ABB=ON  OHLSSON M?/AU
L133         307 SEA FILE=USPATFULL ABB=ON  PETERSON C?/AU
L134           1 SEA FILE=USPATFULL ABB=ON  RINGNER M?/AU
L143           1 SEA FILE=USPATFULL ABB=ON  L131 AND L132 AND L133 AND L134

=> s l142 or l143

L158          1 L142 OR L143

=> dup rem l118,l155,l157,l156,l158
FILE 'PASCAL' ENTERED AT 12:04:07 ON 15 AUG 2003
Any reproduction or dissemination in part or in full,
by means of any process and on any support whatsoever
is prohibited without the prior written agreement of INIST-CNRS.
COPYRIGHT (C) 2003 INIST-CNRS. All rights reserved.

FILE 'BIOSIS' ENTERED AT 12:04:07 ON 15 AUG 2003
COPYRIGHT (C) 2003 BIOLOGICAL ABSTRACTS INC.(R)

FILE 'SCISEARCH' ENTERED AT 12:04:07 ON 15 AUG 2003
COPYRIGHT 2003 THOMSON ISI

FILE 'CAPLUS' ENTERED AT 12:04:07 ON 15 AUG 2003
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
COPYRIGHT (C) 2003 AMERICAN CHEMICAL SOCIETY (ACS)

FILE 'EMBASE' ENTERED AT 12:04:07 ON 15 AUG 2003
COPYRIGHT (C) 2003 Elsevier Science B.V.  All rights reserved.

FILE 'WPIDS' ENTERED AT 12:04:07 ON 15 AUG 2003
COPYRIGHT (C) 2003 THOMSON DERWENT

FILE 'USPATFULL' ENTERED AT 12:04:07 ON 15 AUG 2003
CA INDEXING COPYRIGHT (C) 2003 AMERICAN CHEMICAL SOCIETY (ACS)
PROCESSING COMPLETED FOR L118
PROCESSING COMPLETED FOR L155
PROCESSING COMPLETED FOR L157
PROCESSING COMPLETED FOR L156
PROCESSING COMPLETED FOR L158
L159            4 DUP REM L118 L155 L157 L156 L158 (5 DUPLICATES REMOVED)
```

ANSWER '1' FROM FILE PASCAL
ANSWERS '2-3' FROM FILE CAPLUS
ANSWER '4' FROM FILE USPATFULL

=> d ibib ab 1-4

L159 ANSWER 1 OF 4 PASCAL DUPLICATE 1

ACCESSION NUMBER:	2002-0555482 PASCAL
TITLE (IN ENGLISH):	A novel approach to local reliability of sequence alignments
AUTHOR:	SCHLOSSHAUER Maximilian; **OHLSSON Mattias**
CORPORATE SOURCE:	Complex Systems Division, Department of Theoretical Physics, University of Lund, Solvegatan 14A, 223 62 Lund, Sweden
SOURCE:	Bioinformatics : (Oxford. Print), (2002), 18(6), 847-854, 18 refs. ISSN: 1367-4803
DOCUMENT TYPE:	Journal
BIBLIOGRAPHIC LEVEL:	Analytic
COUNTRY:	United Kingdom
LANGUAGE:	English
AVAILABILITY:	INIST-21331

AB Motivation: The pairwise alignment of biological sequences obtained from an **algorithm** will in general contain both correct and incorrect parts. Hence, to allow for a valid interpretation of the alignment, the local trustworthiness of the alignment has to be quantified. Results: We present a novel approach that attributes a reliability index to every pair of residues, including gapped regions, in the optimal **alignment** of two **protein** sequences. The method is based on a fuzzy recast of the dynamic programming **algorithm** for sequence alignment in terms of **mean field** annealing.
An extensive evaluation with structural reference alignments not only shows that the probability for a pair of residues to be correctly aligned grows consistently with increasing reliability index, but moreover demonstrates that the value of the reliability index can directly be translated into an estimate of the probability for a correct alignment.

L159 ANSWER 2 OF 4 CAPLUS DUPLICATE 2

ACCESSION NUMBER:	2001:748109 CAPLUS
DOCUMENT NUMBER:	135:285367
TITLE:	A method for protein structure alignment
INVENTOR(S):	**Blankenbecler, Richard; Ohlsson, Mattias; Peterson, Carsten; Ringner, Markus**
PATENT ASSIGNEE(S):	Board of Trustees of the Leland Stanford Junior University, USA
SOURCE:	PCT Int. Appl., 35 pp. CODEN: PIXXD2
DOCUMENT TYPE:	Patent
LANGUAGE:	English
FAMILY ACC. NUM. COUNT:	1

PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2001075436	A1	20011011	WO 2001-US10675	20010402
W: CA				
RW: AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR				
US 2002111781	A1	20020815	US 2001-825441	20010402
EP 1272840	A1	20030108	EP 2001-924605	20010402
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT,				

IE, SI, LT, LV, FI, RO, MK, CY, AL, TR
PRIORITY APPLN. INFO.: US 2000-194203P P 20000403
WO 2001-US10675 W 20010402

AB This invention provides a method for protein structure alignment. More particularly, the present invention provides a method for identification, classification and prediction of protein structures. The present invention involves two key ingredients. First, an energy or cost function formulation of the problem simultaneously in terms of binary (Potts) assignment variables and real-valued at. coordinates. Second, a minimization of the energy or cost function by an iterative method, where in each iteration (1) a mean field method is employed for the assignment variables and (2) exact rotation and/or translation of at. coordinates is performed, weighted with the corresponding assignment variables.

REFERENCE COUNT: 5 THERE ARE 5 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L159 ANSWER 3 OF 4 CAPLUS COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER: 1996:553155 CAPLUS
DOCUMENT NUMBER: 125:215105
TITLE: Evidence for **nonrandom** hydrophobicity structures in protein chains
AUTHOR(S): Irbaeck, Anders; **Peterson, Carsten;** Potthast, Frank
CORPORATE SOURCE: Dep. Theoretical Physics, Univ. Lund, Lund, S-223 62, Swed.
SOURCE: Proceedings of the National Academy of Sciences of the United States of America (1996), 93(18), 9533-9538
CODEN: PNASA6; ISSN: 0027-8424
PUBLISHER: National Academy of Sciences
DOCUMENT TYPE: Journal
LANGUAGE: English

AB The question of whether proteins originate from random sequences of amino acids is addressed. A statistical anal. is performed in terms of blocked and random walk values formed by binary hydrophobic assignments of the amino acids along the protein chains. Theor. expectations of these variables from random distributions of hydrophobicities are compared with those obtained from functional proteins. The results, which are based upon proteins in the SWISS-PROT data base, convincingly show that the amino acid sequences in proteins differ from what is expected from random sequences in a statistically significant way. By performing Fourier transforms on the random walks, one obtains addnl. evidence for nonrandomness of the distributions. The authors have also analyzed results from a synthetic model contg. only two amino acid types, hydrophobic and hydrophilic. With reasonable criteria on good folding properties in terms of thermodynamical and kinetic behavior, sequences that fold well are isolated. Performing the same statistical anal. on the sequences that fold well indicates similar deviations from randomness as for the functional proteins. The deviations from randomness can be interpreted as originating from anticorrelations in terms of an Ising spin model for the hydrophobicities. The authors' results, which differ from some previous investigations using other methods, might have impact on how permissive with respect to sequence specificity the protein folding process is-only sequences with nonrandom hydrophobicity distributions fold well. Other distributions give rise to energy landscapes with poor folding properties and hence did not survive the evolution.

L159 ANSWER 4 OF 4 USPATFULL on STN
ACCESSION NUMBER: 2002:207077 USPATFULL
TITLE: Method for **protein structure alignment**
INVENTOR(S): **Blankenbecler, Richard**, Stanford, CA, UNITED STATES
Ohlsson, Mattias, Lund, SWEDEN

Peterson, Carsten, Lund, SWEDEN
Ringner, Markus, Lund, SWEDEN

	NUMBER	KIND	DATE	
PATENT INFORMATION:	US 2002111781	A1	20020815	
APPLICATION INFO.:	US 2001-825441	A1	20010402	(9)

	NUMBER	DATE
PRIORITY INFORMATION:	US 2000-194203P	20000403 (60)

DOCUMENT TYPE: Utility
FILE SEGMENT: APPLICATION
LEGAL REPRESENTATIVE: MAREK ALBOSZTA, LUMEN INTELLECTUAL PROPERTY SERVICES, 45 CABOT AVENUE, SUITE 110, SANTA CLARA, CA, 95051
NUMBER OF CLAIMS: 18
EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 4 Drawing Page(s)
LINE COUNT: 749
CAS INDEXING IS AVAILABLE FOR THIS PATENT.

AB This invention provides a method for **protein structure alignment**. More particularly, the present invention provides a method for identification, classification and prediction of **protein structures**. The present invention involves two key ingredients. First, an energy or cost function formulation of the problem simultaneously in terms of binary (Potts) assignment variables and real-valued atomic coordinates. Second, a minimization of the energy or cost function by an iterative method, where in each iteration (1) a **mean field** method is employed for the assignment variables and (2) exact rotation and/or translation of atomic coordinates is performed, weighted with the corresponding assignment variables.

```
=> fil capl; d que l17; d que l19; d que l21; d que l25; d que l26
FILE 'CAPLUS' ENTERED AT 12:11:32 ON 15 AUG 2003
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
COPYRIGHT (C) 2003 AMERICAN CHEMICAL SOCIETY (ACS)

Copyright of the articles to which records in this database refer is
held by the publishers listed in the PUBLISHER (PB) field (available
for records published or updated in Chemical Abstracts after December
26, 1996), unless otherwise indicated in the original publications.
The CA Lexicon is the copyrighted intellectual property of the
American Chemical Society and is provided to assist you in searching
databases on STN.  Any dissemination, distribution, copying, or storing
of this information, without the prior written consent of CAS, is
strictly prohibited.

FILE COVERS 1907 - 15 Aug 2003  VOL 139 ISS 8
FILE LAST UPDATED: 14 Aug 2003  (20030814/ED)

  This file contains CAS Registry Numbers for easy and accurate
  substance identification.


L14            3 SEA FILE=CAPLUS ABB=ON  BINARY ASSIGNMENT
L16       265249 SEA FILE=CAPLUS ABB=ON  VARIABLE#
L17            1 SEA FILE=CAPLUS ABB=ON  L14 AND L16


L9        739531 SEA FILE=CAPLUS ABB=ON  PROTEINS/CW
L15        19062 SEA FILE=CAPLUS ABB=ON  MEAN FIELD
L16       265249 SEA FILE=CAPLUS ABB=ON  VARIABLE#
L18           29 SEA FILE=CAPLUS ABB=ON  L15 (5A)L16
L19            1 SEA FILE=CAPLUS ABB=ON  L18 AND L9


L7         34837 SEA FILE=CAPLUS ABB=ON  ALGORITHM/CT
L9        739531 SEA FILE=CAPLUS ABB=ON  PROTEINS/CW
L10        22499 SEA FILE=CAPLUS ABB=ON  L9(L)(STRUCTURE? OR ALIGN?)
L15        19062 SEA FILE=CAPLUS ABB=ON  MEAN FIELD
L21            2 SEA FILE=CAPLUS ABB=ON  L15 AND (L10 AND L7)


L9        739531 SEA FILE=CAPLUS ABB=ON  PROTEINS/CW
L10        22499 SEA FILE=CAPLUS ABB=ON  L9(L)(STRUCTURE? OR ALIGN?)
L15        19062 SEA FILE=CAPLUS ABB=ON  MEAN FIELD
L22         9670 SEA FILE=CAPLUS ABB=ON  ENERGY FUNCTION?
L25            4 SEA FILE=CAPLUS ABB=ON  L10 AND L22 AND L15


L6          1008 SEA FILE=CAPLUS ABB=ON  ATOM?(L)DISTANCE#/OBI
L9        739531 SEA FILE=CAPLUS ABB=ON  PROTEINS/CW
L10        22499 SEA FILE=CAPLUS ABB=ON  L9(L)(STRUCTURE? OR ALIGN?)
L26            5 SEA FILE=CAPLUS ABB=ON  L6 AND L10

=> d que l30; d que l31
```

text search

THIS PAGE BLANK (USPTO)

intentionally
blank

```
L9       739531 SEA FILE=CAPLUS ABB=ON  PROTEINS/CW
L15       19062 SEA FILE=CAPLUS ABB=ON  MEAN FIELD
L16      265249 SEA FILE=CAPLUS ABB=ON  VARIABLE#
L18          29 SEA FILE=CAPLUS ABB=ON  L15 (5A)L16
L27      102247 SEA FILE=CAPLUS ABB=ON  "CONFORMATION AND CONFORMERS"/CT
L30           1 SEA FILE=CAPLUS ABB=ON  L18 AND L9 AND L27


L6         1008 SEA FILE=CAPLUS ABB=ON  ATOM?(L)DISTANCE#/OBI
L9       739531 SEA FILE=CAPLUS ABB=ON  PROTEINS/CW
L27      102247 SEA FILE=CAPLUS ABB=ON  "CONFORMATION AND CONFORMERS"/CT
L31           3 SEA FILE=CAPLUS ABB=ON  L9 AND L27 AND L6

=> s (l17 or l19 or l21 or l25 or l26 or l30 or l31) not (l12-l13)

L160          13 (L17 OR L19 OR L21 OR L25 OR L26 OR L30 OR L31) NOT ((L12 OR
                L13))

=> fil wpids; d que l44; d que l48; d que l50

FILE 'WPIDS' ENTERED AT 12:11:34 ON 15 AUG 2003
COPYRIGHT (C) 2003 THOMSON DERWENT

FILE LAST UPDATED:            13 AUG 2003       <20030813/UP>
MOST RECENT DERWENT UPDATE:      200352          <200352/DW>
DERWENT WORLD PATENTS INDEX SUBSCRIBER FILE, COVERS 1963 TO DATE

>>> NEW WEEKLY SDI FREQUENCY AVAILABLE --> see NEWS  <<<

>>> PATENT IMAGES AVAILABLE FOR PRINT AND DISPLAY <<<

>>> FOR DETAILS OF THE PATENTS COVERED IN CURRENT UPDATES,
    SEE http://www.derwent.com/dwpi/updates/dwpicov/index.html <<<

>>> FOR A COPY OF THE DERWENT WORLD PATENTS INDEX STN USER GUIDE,
    PLEASE VISIT:
 http://www.stn-international.de/training_center/patents/stn_guide.pdf <<<

>>> FOR INFORMATION ON ALL DERWENT WORLD PATENTS INDEX USER
    GUIDES, PLEASE VISIT:
    http://www.derwent.com/userguides/dwpi_guide.html <<<


L37          31 SEA FILE=WPIDS ABB=ON  MEAN FIELD
L38           2 SEA FILE=WPIDS ABB=ON  BINARY ASSIGNMENT#
L39      114374 SEA FILE=WPIDS ABB=ON  PROTEIN#
L40        2132 SEA FILE=WPIDS ABB=ON  L39(5A)(STRUCTURE# OR CONFORM? OR
                ALIGN?)
L44           4 SEA FILE=WPIDS ABB=ON  L40 AND (L37 OR L38)


L39      114374 SEA FILE=WPIDS ABB=ON  PROTEIN#
L40        2132 SEA FILE=WPIDS ABB=ON  L39(5A)(STRUCTURE# OR CONFORM? OR
                ALIGN?)
L41         132 SEA FILE=WPIDS ABB=ON  ENERGY FUNCTION#
L42         220 SEA FILE=WPIDS ABB=ON  ATOM?(2A)DISTANCE#
L48           1 SEA FILE=WPIDS ABB=ON  L40 AND L41 AND L42
```

previously printed w/ inventors search

```
L39       114374 SEA FILE=WPIDS ABB=ON  PROTEIN#
L40         2132 SEA FILE=WPIDS ABB=ON  L39(5A)(STRUCTURE# OR CONFORM? OR
                 ALIGN?)
L41          132 SEA FILE=WPIDS ABB=ON  ENERGY FUNCTION#
L42          220 SEA FILE=WPIDS ABB=ON  ATOM?(2A)DISTANCE#
L49       469670 SEA FILE=WPIDS ABB=ON  ALGORITH? OR COMPUT?
L50            9 SEA FILE=WPIDS ABB=ON  L40 AND (L41 OR L42) AND L49
```

```
=> s (144 or 148 or 150) not (136 or 143)
```

previously printed

```
L161        10 (L44 OR L48 OR L50) NOT (L36 OR L43)

=> fil medl; d que 168; d que 171; d que 175; d que 180

FILE 'MEDLINE' ENTERED AT 12:11:36 ON 15 AUG 2003

 FILE LAST UPDATED: 14 AUG 2003 (20030814/UP).  FILE COVERS 1958 TO DATE.

 On April 13, 2003, MEDLINE was reloaded.  See HELP RLOAD for details.

 MEDLINE thesauri in the /CN, /CT, and /MN fields incorporate the
 MeSH 2003 vocabulary.  See http://www.nlm.nih.gov/mesh/changes2003.html
 for a description on changes.

 This file contains CAS Registry Numbers for easy and accurate
 substance identification.


L58       130531 SEA FILE=MEDLINE ABB=ON  PROTEIN CONFORMATION+NT/CT
L60            3 SEA FILE=MEDLINE ABB=ON  BINARY ASSIGNMENT#
L68            0 SEA FILE=MEDLINE ABB=ON  L58 AND L60


L58       130531 SEA FILE=MEDLINE ABB=ON  PROTEIN CONFORMATION+NT/CT
L59          226 SEA FILE=MEDLINE ABB=ON  ATOM?(2A)DISTANCE#
L61          734 SEA FILE=MEDLINE ABB=ON  MEAN FIELD
L62          580 SEA FILE=MEDLINE ABB=ON  ENERGY FUNCTION#
L71            6 SEA FILE=MEDLINE ABB=ON  L58 AND L59 AND (L61 OR L62)


L58       130531 SEA FILE=MEDLINE ABB=ON  PROTEIN CONFORMATION+NT/CT
L61          734 SEA FILE=MEDLINE ABB=ON  MEAN FIELD
L62          580 SEA FILE=MEDLINE ABB=ON  ENERGY FUNCTION#
L75            5 SEA FILE=MEDLINE ABB=ON  L61 AND L62 AND L58


L58       130531 SEA FILE=MEDLINE ABB=ON  PROTEIN CONFORMATION+NT/CT
L59          226 SEA FILE=MEDLINE ABB=ON  ATOM?(2A)DISTANCE#
L61          734 SEA FILE=MEDLINE ABB=ON  MEAN FIELD
L62          580 SEA FILE=MEDLINE ABB=ON  ENERGY FUNCTION#
L64        36956 SEA FILE=MEDLINE ABB=ON  ALGORITHMS/CT
L65        48459 SEA FILE=MEDLINE ABB=ON  SOFTWARE+NT/CT
L76          139 SEA FILE=MEDLINE ABB=ON  L58/MAJ AND (L59 OR L61 OR L62)
L80            9 SEA FILE=MEDLINE ABB=ON  L76 AND L65 AND L64

=> s (171 or 175 or 180)
```

```
L162          19 (L71 OR L75 OR L80)

=> fil embase

FILE 'EMBASE' ENTERED AT 12:11:38 ON 15 AUG 2003
```


```
 FILE COVERS 1974 TO 14 Aug 2003 (20030814/ED)

 EMBASE has been reloaded. Enter HELP RLOAD for details.

 This file contains CAS Registry Numbers for easy and accurate
 substance identification.

=> d que l94; d que l95; d que l99; d que l105

L81       219472 SEA FILE=EMBASE ABB=ON  PROTEIN STRUCTURE+NT/CT
L86          303 SEA FILE=EMBASE ABB=ON  ATOM?(3A)DISTANCE#
L87            2 SEA FILE=EMBASE ABB=ON  BINARY ASSIGNMENT#
L88          276 SEA FILE=EMBASE ABB=ON  MEAN FIELD#
L89          487 SEA FILE=EMBASE ABB=ON  ENERGY FUNCTION#
L94            4 SEA FILE=EMBASE ABB=ON  L81 AND L86 AND (L87 OR L88 OR L89)


L81       219472 SEA FILE=EMBASE ABB=ON  PROTEIN STRUCTURE+NT/CT
L87            2 SEA FILE=EMBASE ABB=ON  BINARY ASSIGNMENT#
L95            1 SEA FILE=EMBASE ABB=ON  L81 AND L87


L81       219472 SEA FILE=EMBASE ABB=ON  PROTEIN STRUCTURE+NT/CT
L88          276 SEA FILE=EMBASE ABB=ON  MEAN FIELD#
L89          487 SEA FILE=EMBASE ABB=ON  ENERGY FUNCTION#
L91        22056 SEA FILE=EMBASE ABB=ON  ALGORITHM/CT
L92        23728 SEA FILE=EMBASE ABB=ON  COMPUTER PROGRAM/CT
L99            2 SEA FILE=EMBASE ABB=ON  L81 AND L88 AND L89 AND (L91 OR L92)


L81       219472 SEA FILE=EMBASE ABB=ON  PROTEIN STRUCTURE+NT/CT
L86          303 SEA FILE=EMBASE ABB=ON  ATOM?(3A)DISTANCE#
L87            2 SEA FILE=EMBASE ABB=ON  BINARY ASSIGNMENT#
L88          276 SEA FILE=EMBASE ABB=ON  MEAN FIELD#
L89          487 SEA FILE=EMBASE ABB=ON  ENERGY FUNCTION#
L91        22056 SEA FILE=EMBASE ABB=ON  ALGORITHM/CT
L92        23728 SEA FILE=EMBASE ABB=ON  COMPUTER PROGRAM/CT
L101       55428 SEA FILE=EMBASE ABB=ON  PROTEIN ANALYSIS/CT
L105           8 SEA FILE=EMBASE ABB=ON  L81/MAJ AND L101 AND ((L86 OR L87 OR
                L88 OR L89)) AND (L91 OR L92)
                                                  previously
=> s (l94 or l95 or l99 or l105) not (l90 or l93)      printed

L163          13 (L94 OR L95 OR L99 OR L105) NOT (L90 OR L93)

=> fil PASCAL, BIOTECHNO, ESBIOBASE, LIFESCI, BIOSIS, TOXCENTER, scisearch

FILE 'PASCAL' ENTERED AT 12:11:40 ON 15 AUG 2003
```

FILE 'BIOTECHNO' ENTERED AT 12:11:40 ON 15 AUG 2003

FILE 'ESBIOBASE' ENTERED AT 12:11:40 ON 15 AUG 2003

FILE 'LIFESCI' ENTERED AT 12:11:40 ON 15 AUG 2003

FILE 'BIOSIS' ENTERED AT 12:11:40 ON 15 AUG 2003

FILE 'TOXCENTER' ENTERED AT 12:11:40 ON 15 AUG 2003

FILE 'SCISEARCH' ENTERED AT 12:11:40 ON 15 AUG 2003


```
=> d que l120; d que l122; d que l128; d que l130

L110      355959 SEA PROTEIN#(5A)(STRUCTUR? OR ALIGN? OR CONFORM?)
L112          10 SEA BINARY ASSIGNMENT#
L120           0 SEA L110 AND L112


L110      355959 SEA PROTEIN#(5A)(STRUCTUR? OR ALIGN? OR CONFORM?)
L111        5034 SEA ATOM?(3A) DISTANCE#
L113       27964 SEA MEAN FIELD#
L114       11650 SEA ENERGY FUNCTION?
L115     1695147 SEA COMPUT?
L116      508361 SEA ALGORITH?
L117      213234 SEA COORDINATE#
L122          23 SEA L110 AND L111 AND (L113 OR L114 OR L117) AND (L115 OR
                 L116)


L110      355959 SEA PROTEIN#(5A)(STRUCTUR? OR ALIGN? OR CONFORM?)
L113       27964 SEA MEAN FIELD#
L114       11650 SEA ENERGY FUNCTION?
L115     1695147 SEA COMPUT?
L116      508361 SEA ALGORITH?
L128           2 SEA L113 AND L114 AND L110 AND L115 AND L116


L110      355959 SEA PROTEIN#(5A)(STRUCTUR? OR ALIGN? OR CONFORM?)
L111        5034 SEA ATOM?(3A) DISTANCE#
L113       27964 SEA MEAN FIELD#
L114       11650 SEA ENERGY FUNCTION?
L115     1695147 SEA COMPUT?
L116      508361 SEA ALGORITH?
L117      213234 SEA COORDINATE#
L130          12 SEA L110 AND (L111 OR L113 OR L114) AND L117 AND L115 AND L116

=> s (l122 or l128 or l130) not l118

L164          33 (L122 OR L128 OR L130) NOT L118
```

previously printed

```
=> fil uspatf; d que l144; d que l147; d que l149

FILE 'USPATFULL' ENTERED AT 12:11:45 ON 15 AUG 2003
CA INDEXING COPYRIGHT (C) 2003 AMERICAN CHEMICAL SOCIETY (ACS)

FILE COVERS 1971 TO PATENT PUBLICATION DATE: 14 Aug 2003 (20030814/PD)
FILE LAST UPDATED: 14 Aug 2003 (20030814/ED)
HIGHEST GRANTED PATENT NUMBER: US6606748
HIGHEST APPLICATION PUBLICATION NUMBER: US2003154532
CA INDEXING IS CURRENT THROUGH 14 Aug 2003 (20030814/UPCA)
ISSUE CLASS FIELDS (/INCL) CURRENT THROUGH: 14 Aug 2003 (20030814/PD)
REVISED CLASS FIELDS (/NCL) LAST RELOADED: Jun 2003
USPTO MANUAL OF CLASSIFICATIONS THESAURUS ISSUE DATE: Jun 2003

>>>  USPAT2 is now available.  USPATFULL contains full text of the     <<<
>>>  original, i.e., the earliest published granted patents or         <<<
>>>  applications.  USPAT2 contains full text of the latest US         <<<
>>>  publications, starting in 2001, for the inventions covered in     <<<
>>>  USPATFULL.  A USPATFULL record contains not only the original     <<<
>>>  published document but also a list of any subsequent              <<<
>>>  publications.  The publication number, patent kind code, and      <<<
>>>  publication date for all the US publications for an invention     <<<
>>>  are displayed in the PI (Patent Information) field of USPATFULL   <<<
>>>  records and may be searched in standard search fields, e.g., /PN, <<<
>>>  /PK, etc.                                                         <<<

>>>  USPATFULL and USPAT2 can be accessed and searched together        <<<
>>>  through the new cluster USPATALL.  Type FILE USPATALL to          <<<
>>>  enter this cluster.                                               <<<
>>>                                                                    <<<
>>>  Use USPATALL when searching terms such as patent assignees,       <<<
>>>  classifications, or claims, that may potentially change from      <<<
>>>  the earliest to the latest publication.                           <<<

This file contains CAS Registry Numbers for easy and accurate
substance identification.


L135        1604 SEA FILE=USPATFULL ABB=ON  PROTEIN#(5A)(STRUCTUR? OR ALIGN? OR
                CONFORM?)/IT,TI,AB,CLM
L137          32 SEA FILE=USPATFULL ABB=ON  BINARY ASSIGNMENT# OR (BINARY
                ASSIGNMENT#)/IT
L144           1 SEA FILE=USPATFULL ABB=ON  L135 AND L137


L135        1604 SEA FILE=USPATFULL ABB=ON  PROTEIN#(5A)(STRUCTUR? OR ALIGN? OR
                CONFORM?)/IT,TI,AB,CLM
L136        2262 SEA FILE=USPATFULL ABB=ON  ATOM?(3A)DISTANCE# OR (ATOM?(3A)DIST
                ANCE#)/IT
L138         228 SEA FILE=USPATFULL ABB=ON  MEAN FIELD# OR (MEAN FIELD#)/IT
L139         954 SEA FILE=USPATFULL ABB=ON  ENERGY FUNCTION? OR (ENERGY
                FUNCTION?)/IT
L147           6 SEA FILE=USPATFULL ABB=ON  L135 AND L136 AND L138 AND L139


L135        1604 SEA FILE=USPATFULL ABB=ON  PROTEIN#(5A)(STRUCTUR? OR ALIGN? OR
                CONFORM?)/IT,TI,AB,CLM
L136        2262 SEA FILE=USPATFULL ABB=ON  ATOM?(3A)DISTANCE# OR (ATOM?(3A)DIST
                ANCE#)/IT
L138         228 SEA FILE=USPATFULL ABB=ON  MEAN FIELD# OR (MEAN FIELD#)/IT
L139         954 SEA FILE=USPATFULL ABB=ON  ENERGY FUNCTION? OR (ENERGY
```

```
               FUNCTION?)/IT
L140     159150 SEA FILE=USPATFULL ABB=ON  ALGORITH? OR ALGORITH?/IT
L141     615026 SEA FILE=USPATFULL ABB=ON  COMPUT? OR COMPUT?/IT
L149          1 SEA FILE=USPATFULL ABB=ON  L135 (P) L136 (P) (L138 OR L139)
               (P) L140 (P) L141
```

previously printed

```
=> s (l144 or l147 or l149) not (l142 or l143)

L165          6 (L144 OR L147 OR L149) NOT (L142 OR L143)

=> dup rem l162,l164,l160,l163,l161,l165
FILE 'MEDLINE' ENTERED AT 12:12:36 ON 15 AUG 2003

FILE 'PASCAL' ENTERED AT 12:12:36 ON 15 AUG 2003
Any reproduction or dissemination in part or in full,
by means of any process and on any support whatsoever
is prohibited without the prior written agreement of INIST-CNRS.
COPYRIGHT (C) 2003 INIST-CNRS. All rights reserved.

FILE 'BIOTECHNO' ENTERED AT 12:12:36 ON 15 AUG 2003
COPYRIGHT (C) 2003 Elsevier Science B.V., Amsterdam. All rights reserved.

FILE 'ESBIOBASE' ENTERED AT 12:12:36 ON 15 AUG 2003
COPYRIGHT (C) 2003 Elsevier Science B.V., Amsterdam. All rights reserved.

FILE 'BIOSIS' ENTERED AT 12:12:36 ON 15 AUG 2003
COPYRIGHT (C) 2003 BIOLOGICAL ABSTRACTS INC.(R)

FILE 'SCISEARCH' ENTERED AT 12:12:36 ON 15 AUG 2003
COPYRIGHT 2003 THOMSON ISI

FILE 'CAPLUS' ENTERED AT 12:12:36 ON 15 AUG 2003
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
COPYRIGHT (C) 2003 AMERICAN CHEMICAL SOCIETY (ACS)

FILE 'EMBASE' ENTERED AT 12:12:36 ON 15 AUG 2003
COPYRIGHT (C) 2003 Elsevier Science B.V.  All rights reserved.

FILE 'WPIDS' ENTERED AT 12:12:36 ON 15 AUG 2003
COPYRIGHT (C) 2003 THOMSON DERWENT

FILE 'USPATFULL' ENTERED AT 12:12:36 ON 15 AUG 2003
CA INDEXING COPYRIGHT (C) 2003 AMERICAN CHEMICAL SOCIETY (ACS)
PROCESSING COMPLETED FOR L162
PROCESSING COMPLETED FOR L164
PROCESSING COMPLETED FOR L160
PROCESSING COMPLETED FOR L163
PROCESSING COMPLETED FOR L161
PROCESSING COMPLETED FOR L165
L166         70 DUP REM L162 L164 L160 L163 L161 L165 (24 DUPLICATES REMOVED)
               ANSWERS '1-19' FROM FILE MEDLINE
               ANSWERS '20-21' FROM FILE PASCAL
               ANSWERS '22-24' FROM FILE BIOTECHNO
               ANSWERS '25-28' FROM FILE BIOSIS
               ANSWERS '29-36' FROM FILE SCISEARCH
               ANSWERS '37-46' FROM FILE CAPLUS
               ANSWERS '47-54' FROM FILE EMBASE
               ANSWERS '55-64' FROM FILE WPIDS
               ANSWERS '65-70' FROM FILE USPATFULL

=> d ibib ab 1-70; fil hom
```

L166 ANSWER 1 OF 70 MEDLINE on STN DUPLICATE 4
ACCESSION NUMBER: 2000500243 MEDLINE
DOCUMENT NUMBER: 20498320 PubMed ID: 11045621
TITLE: Modeling of loops in protein structures.
AUTHOR: Fiser A; Do R K; Sali A
CORPORATE SOURCE: Laboratory of Molecular Biophysics, Pels Family Center for Biochemistry and Structural Biology, The Rockefeller University, New York, New York 10021, USA.. sali@rockefeller.edu
CONTRACT NUMBER: GM 54762 (NIGMS)
SOURCE: PROTEIN SCIENCE, (2000 Sep) 9 (9) 1753-73. Journal code: 9211750. ISSN: 0961-8368.
PUB. COUNTRY: United States
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 200102
ENTRY DATE: Entered STN: 20010322
Last Updated on STN: 20010322
Entered Medline: 20010201

AB Comparative protein structure prediction is limited mostly by the errors in alignment and loop modeling. We describe here a new automated modeling technique that significantly improves the accuracy of loop predictions in protein structures. The positions of all nonhydrogen atoms of the loop are optimized in a fixed environment with respect to a pseudo **energy function**. The energy is a sum of many spatial restraints that include the bond length, bond angle, and improper dihedral angle terms from the CHARMM-22 force field, statistical preferences for the main-chain and side-chain dihedral angles, and statistical preferences for nonbonded atomic contacts that depend on the two **atom** types, their **distance** through space, and separation in sequence. The **energy function** is optimized with the method of conjugate gradients combined with molecular dynamics and simulated annealing. Typically, the predicted loop conformation corresponds to the lowest energy conformation among 500 independent optimizations. Predictions were made for 40 loops of known structure at each length from 1 to 14 residues. The accuracy of loop predictions is evaluated as a function of thoroughness of conformational sampling, loop length, and structural properties of native loops. When accuracy is measured by local superposition of the model on the native loop, 100, 90, and 30% of 4-, 8-, and 12-residue loop predictions, respectively, had <2 A RMSD error for the mainchain N, C(alpha), C, and O atoms; the average accuracies were 0.59 +/- 0.05, 1.16 +/- 0.10, and 2.61 +/- 0.16 A, respectively. To simulate real comparative modeling problems, the method was also evaluated by predicting loops of known structure in only approximately correct environments with errors typical of comparative modeling without misalignment. When the RMSD distortion of the main-chain stem atoms is 2.5 A, the average loop prediction error increased by 180, 25, and 3% for 4-, 8-, and 12-residue loops, respectively. The accuracy of the lowest energy prediction for a given loop can be estimated from the structural variability among a number of low energy predictions. The relative value of the present method is gauged by (1) comparing it with one of the most successful previously described methods, and (2) describing its accuracy in recent blind predictions of protein structure. Finally, it is shown that the average accuracy of prediction is limited primarily by the accuracy of the **energy function** rather than by the extent of conformational sampling.

L166 ANSWER 2 OF 70 MEDLINE on STN DUPLICATE 5
ACCESSION NUMBER: 2000484127 MEDLINE
DOCUMENT NUMBER: 20368715 PubMed ID: 10906342
TITLE: Identifying sequence-structure pairs undetected by sequence

alignments.
AUTHOR: Miyazawa S; Jernigan R L
CORPORATE SOURCE: Faculty of Technology, Gunma University, Kiryu, Gunma 376, Japan and Room B-116, Bldg 12B, MSC 5677, Laboratory of Experimental and Computational Biology, DBS, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892-5677,USA.
SOURCE: PROTEIN ENGINEERING, (2000 Jul) 13 (7) 459-75.
Journal code: 8801484. ISSN: 0269-2139.
PUB. COUNTRY: ENGLAND: United Kingdom
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 200010
ENTRY DATE: Entered STN: 20001019
Last Updated on STN: 20001019
Entered Medline: 20001012

AB We examine how effectively simple potential functions previously developed can identify compatibilities between sequences and structures of proteins for database searches. The potential function consists of pairwise contact energies, repulsive packing potentials of residues for overly dense arrangement and short-range potentials for secondary structures, all of which were estimated from statistical preferences observed in known protein structures. Each potential energy term was modified to represent compatibilities between sequences and structures for globular proteins. Pairwise contact interactions in a sequence-structure alignment are evaluated in a **mean field** approximation on the basis of probabilities of site pairs to be aligned. Gap penalties are assumed to be proportional to the number of contacts at each residue position, and as a result gaps will be more frequently placed on protein surfaces than in cores. In addition to minimum energy alignments, we use probability alignments made by successively aligning site pairs in order by pairwise alignment probabilities. The results show that the present **energy function** and alignment method can detect well both folds compatible with a given sequence and, inversely, sequences compatible with a given fold, and yield mostly similar alignments for these two types of sequence and structure pairs. Probability alignments consisting of most reliable site pairs only can yield extremely small root mean square deviations, and including less reliable pairs increases the deviations. Also, it is observed that secondary structure potentials are usefully complementary to yield improved alignments with this method. Remarkably, by this method some individual sequence-structure pairs are detected having only 5-20% sequence identity.

L166 ANSWER 3 OF 70 MEDLINE on STN DUPLICATE 6
ACCESSION NUMBER: 2000095849 MEDLINE
DOCUMENT NUMBER: 20095849 PubMed ID: 10631988
TITLE: Protein structure determination using a database of interatomic distance probabilities.
AUTHOR: Wall M E; Subramaniam S; Phillips G N Jr
CORPORATE SOURCE: Department of Biochemistry and Cell Biology and Keck Center for Computational Biology, Rice University, Houston, Texas 77005-1892, USA.. mewall@lanl.gov
CONTRACT NUMBER: AR 40252 (NIAMS)
SOURCE: PROTEIN SCIENCE, (1999 Dec) 8 (12) 2720-7.
Journal code: 9211750. ISSN: 0961-8368.
PUB. COUNTRY: United States
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
OTHER SOURCE: PDB-1CML; PDB-1MBD; PDB-1OSA
ENTRY MONTH: 200002
ENTRY DATE: Entered STN: 20000229

Last Updated on STN: 20000229
Entered Medline: 20000214

AB The accelerated pace of genomic sequencing has increased the demand for structural models of gene products. Improved quantitative methods are needed to study the many systems (e.g., macromolecular assemblies) for which data are scarce. Here, we describe a new molecular dynamics method for protein structure determination and molecular modeling. An **energy function**, or database potential, is derived from distributions of interatomic distances obtained from a database of known structures. X-ray crystal structures are refined by molecular dynamics with the new **energy function** replacing the Van der Waals potential. Compared to standard methods, this method improved the **atomic** positions, interatomic **distances**, and side-chain dihedral angles of structures randomized to mimic the early stages of refinement. The greatest enhancement in side-chain placement was observed for groups that are characteristically buried. More accurate calculated model phases will follow from improved interatomic distances. Details usually seen only in high-resolution refinements were improved, as is shown by an R-factor analysis. The improvements were greatest when refinements were carried out using X-ray data truncated at 3.5 A. The database potential should therefore be a valuable tool for determining X-ray structures, especially when only low-resolution data are available.

L166 ANSWER 4 OF 70 MEDLINE on STN DUPLICATE 7

ACCESSION NUMBER:	1999309434 MEDLINE
DOCUMENT NUMBER:	99309434 PubMed ID: 10380336
TITLE:	Improvement of side-chain modeling in proteins with the self-consistent **mean field** theory method based on an analysis of the factors influencing prediction.
AUTHOR:	Mendes J; Soares C M; Carrondo M A
CORPORATE SOURCE:	Instituto de Tecnologia Quimica e Biologica, Universidade Nova de Lisboa, Oeiras, Portugal.
SOURCE:	BIOPOLYMERS, (1999 Aug) 50 (2) 111-31. Journal code: 0372525. ISSN: 0006-3525.
PUB. COUNTRY:	United States
DOCUMENT TYPE:	Journal; Article; (JOURNAL ARTICLE)
LANGUAGE:	English
FILE SEGMENT:	Priority Journals
ENTRY MONTH:	199907
ENTRY DATE:	Entered STN: 19990806 Last Updated on STN: 19990806 Entered Medline: 19990726

AB With the objective of improving side-chain conformation prediction, we have analyzed the influence of various factors on prediction by the Self-Consistent **Mean Field** Theory method, applied to a set of high resolution x-ray protein structure models. These factors may be classed as variations in the **mean field** optimization protocol, variations in the potential **energy function**, and variations in rotamer library completeness. We have developed an optimization protocol that consistently reached lower **mean field** conformational free energies than two other protocols. This protocol led to an important improvement in prediction. We observed a major improvement in prediction with two more detailed van der Waals parameter sets, which we found to be due mainly to the introduction of scaling of 1-4 interactions. In a comparison of two knowledge-based rotamer libraries of considerably different size, we observed an unexpected decrease in prediction with an increase in library completeness. However, when we introduced a torsion potential term in the potential **energy function**, we found an important increase in average prediction and in the prediction of almost all residue types with a more complete rotamer set. The two knowledge-based rotamer libraries now became equivalent in terms of average prediction. The

results we obtained in an analysis of the effect of the introduction of an additional electrostatic term in the potential **energy function** were largely inconclusive. However, we found a small increase in average prediction for an electrostatic potential term with a fixed dielectric constant of 15. The combined effect of all the factors we analyzed in this study resulted in average prediction accuracies of 79.9% for X1, 68.1% for X1 + 2, and 1.590 A for global rms deviation (RMSD); the corresponding values for core residues were 88.2%, 78.6%, and 1.171 A. These values represent improvements in average prediction of 6.5% for X1, 9.1% for X1 + 2, and 0.163 A for global RMSD over the original conditions; the corresponding improvements in the core were 5.9%, 9.0%, and 0.180 A, respectively.

L166 ANSWER 5 OF 70 MEDLINE on STN DUPLICATE 10

ACCESSION NUMBER: 1998179564 MEDLINE
DOCUMENT NUMBER: 98179564 PubMed ID: 9514726
TITLE: Rapid refinement of protein interfaces incorporating solvation: application to the docking problem.
AUTHOR: Jackson R M; Gabb H A; Sternberg M J
CORPORATE SOURCE: Biomolecular Modeling Laboratory, Imperial Cancer Research Fund, London, UK.
SOURCE: JOURNAL OF MOLECULAR BIOLOGY, (1998 Feb 13) 276 (1) 265-85. Journal code: 2985088R. ISSN: 0022-2836.
PUB. COUNTRY: ENGLAND: United Kingdom
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 199804
ENTRY DATE: Entered STN: 19980416
Last Updated on STN: 19980416
Entered Medline: 19980407

AB A computationally tractable strategy has been developed to refine protein-protein interfaces that models the effects of side-chain conformational change, solvation and limited rigid-body movement of the subunits. The proteins are described at the atomic level by a multiple copy representation of side-chains modelled according to a rotamer library on a fixed peptide backbone. The surrounding solvent environment is described by "soft" sphere Langevin dipoles for water that interact with the protein via electrostatic, van der Waals and field-dependent hydrophobic terms. Energy refinement is based on a two-step process in which (1) a probability-based conformational matrix of the protein side-chains is refined iteratively by a **mean field** method. A side-chain interacts with the protein backbone and the probability-weighted average of the surrounding protein side-chains and solvent molecules. The resultant protein conformations then undergo (2) rigid-body energy minimization to relax the protein interface. Steps (1) and (2) are repeated until convergence of the interaction energy. The influence of refinement on side-chain conformation starting from unbound conformations found improvement in the RMSD of side-chains in the interface of protease-inhibitor complexes, and shows that the method leads to an improvement in interface geometry. In terms of discriminating between docked structures, the refinement was applied to two classes of protein-protein complex: five protease-protein inhibitor and four antibody-antigen complexes. A large number of putative docked complexes have already been generated for the test systems using our rigid-body docking program, FTDOCK. They include geometries that closely resemble the crystal complex, and therefore act as a test for the refinement procedure. In the protease-inhibitors, geometries that resemble the crystal complex are ranked in the top four solutions for four out of five systems when solvation is included in the **energy function**, against a background of between 26 and 364 complexes in the data set. The results for the antibody-antigen complexes are not as encouraging, with only two of the four systems showing discrimination. It

would appear that these results reflect the somewhat different binding mechanism dominant in the two types of protein-protein complex. Binding in the protease-inhibitors appears to be "lock and key" in nature. The fixed backbone and mobile side-chain representation provide a good model for binding. Movements in the backbone geometry of antigens on binding represent an "induced-fit" and provides more of a challenge for the model. Given the limitations of the conformational sampling, the ability of the **energy function** to discriminate between native and non-native states is encouraging. Development of the approach to include greater conformational sampling could lead to a more general solution to the protein docking problem.

L166 ANSWER 6 OF 70 MEDLINE on STN DUPLICATE 11
ACCESSION NUMBER: 1998147461 MEDLINE
DOCUMENT NUMBER: 98147461 PubMed ID: 9488137
TITLE: Recognition of protein structure on coarse lattices with residue-residue **energy functions**.
AUTHOR: Reva B A; Finkelstein A V; Sanner M; Olson A J; Skolnick J
CORPORATE SOURCE: Department of Molecular Biology, The Scripps Research Institute, CA 92037, USA.
CONTRACT NUMBER: GM48835 (NIGMS)
P01GM38794 (NIGMS)
TW00546 (FIC)
SOURCE: PROTEIN ENGINEERING, (1997 Oct) 10 (10) 1123-30.
Journal code: 8801484. ISSN: 0269-2139.
PUB. COUNTRY: ENGLAND: United Kingdom
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 199803
ENTRY DATE: Entered STN: 19980410
Last Updated on STN: 20000303
Entered Medline: 19980331

AB We suggest and test potentials for the modeling of protein structure on coarse lattices. The coarser the lattice, the more complete and faster is the exploration of the conformational space of a molecule. However, there are inevitable energy errors in lattice modeling caused by distortions in distances between interacting residues; the coarser the lattice, the larger are the energy errors. It is generally believed that an improvement in the accuracy of lattice modelling can be achieved only by reducing the lattice spacing. We reduce the errors on coarse lattices with lattice-adapted potentials. Two methods are used: in the first approach, 'lattice-derived' potentials are obtained directly from a database of lattice models of protein structure; in the second approach, we derive 'lattice-adjusted' potentials using our previously developed method of statistical adjustment of the 'off-lattice' **energy functions** for lattices. The derivation of off-lattice Calpha **atom**-based **distance**-dependent pairwise potentials has been reported previously. The accuracy of 'lattice-derived', 'lattice-adjusted' and 'off-lattice' potentials is estimated in threading tests. It is shown that 'lattice-derived' and 'lattice-adjusted' potentials give virtually the same accuracy and ensure reasonable protein fold recognition on the coarsest considered lattice (spacing 3.8 A), however, the 'off-lattice' potentials, which efficiently recognize off-lattice folds, do not work on this lattice, mainly because of the errors in short-range interactions between neighboring residues.

L166 ANSWER 7 OF 70 MEDLINE on STN DUPLICATE 12
ACCESSION NUMBER: 96097375 MEDLINE
DOCUMENT NUMBER: 96097375 PubMed ID: 8539250
TITLE: Structure and internal dynamics of the bovine pancreatic trypsin inhibitor in aqueous solution from long-time molecular dynamics simulations.

AUTHOR: Brunne R M; Berndt K D; Guntert P; Wuthrich K; van Gunsteren W F
CORPORATE SOURCE: Laboratorium fur Physikalische Chemie, ETH Zentrum, Zurich, Switzerland.
SOURCE: PROTEINS, (1995 Sep) 23 (1) 49-62.
Journal code: 8700181. ISSN: 0887-3585.
PUB. COUNTRY: United States
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 199602
ENTRY DATE: Entered STN: 19960221
Last Updated on STN: 19960221
Entered Medline: 19960208

AB Structural and dynamic properties of bovine pancreatic trypsin inhibitor (BPTI) in aqueous solution are investigated using two molecular dynamics (MD) simulations: one of 1.4 ns length and one of 0.8 ns length in which **atom-atom distance** bounds derived from NMR spectroscopy are included in the potential **energy function** to make the trajectory satisfy these experimental data more closely. The simulated properties of BPTI are compared with crystal and solution structures of BPTI, and found to be in agreement with the available experimental data. The best agreement with experiment was obtained when **atom-atom distance** restraints were applied in a time-averaged manner in the simulation. The polypeptide segments found to be most flexible in the MD simulations coincide closely with those showing differences between the crystal and solution structures of BPTI.

L166 ANSWER 8 OF 70 MEDLINE on STN DUPLICATE 13
ACCESSION NUMBER: 91336385 MEDLINE
DOCUMENT NUMBER: 91336385 PubMed ID: 1872378
TITLE: Constrained optimization and protein structure determination.
AUTHOR: Eisenfeld J; Vajda S; Sugar I; DeLisi C
CORPORATE SOURCE: Department of Biomathematical Sciences, Mount Sinai School of Medicine, New York, New York 10029.
CONTRACT NUMBER: RO1-AI-27471-01 (NIAID)
SOURCE: AMERICAN JOURNAL OF PHYSIOLOGY, (1991 Aug) 261 (2 Pt 1) C376-86.
Journal code: 0370511. ISSN: 0002-9513.
PUB. COUNTRY: United States
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 199109
ENTRY DATE: Entered STN: 19911006
Last Updated on STN: 19911006
Entered Medline: 19910919

AB Energy minimization is one of the main approaches to the computational determination of macromolecular structure. Due to the approximations in the empirical free-**energy functions** and due to the computational difficulties in locating their global minima, the problem is at present intractable when the only information available is the sequence of subunits forming the molecule. A less-demanding problem in terms of both physics and mathematics is constrained optimization, which uses additional but incomplete experimental information such as **distances** between certain **atoms**. This paper reviews methods for generating molecular structure using bond lengths and angles as variables and shows how the structure can be fully specified in terms of local geometry. The analysis permits precise statements to be made about the minimum set of distances that specify a unique structure without recourse to energy minimization. We then discuss the complementary

situation, i.e., structure prediction with energy minimization based only on sequence information. Finally, we show how distance constraints can be incorporated into energy minimization methods.

L166 ANSWER 9 OF 70 MEDLINE on STN DUPLICATE 14
ACCESSION NUMBER: 85264800 MEDLINE
DOCUMENT NUMBER: 85264800 PubMed ID: 3894675
TITLE: A molecular dynamics study of the C-terminal fragment of the L7/L12 ribosomal protein. Secondary structure motion in a 150 picosecond trajectory.
AUTHOR: Aqvist J; van Gunsteren W F; Leijonmarck M; Tapia O
SOURCE: JOURNAL OF MOLECULAR BIOLOGY, (1985 Jun 5) 183 (3) 461-77.
Journal code: 2985088R. ISSN: 0022-2836.
PUB. COUNTRY: ENGLAND: United Kingdom
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 198509
ENTRY DATE: Entered STN: 19900320
Last Updated on STN: 19900320
Entered Medline: 19850925

AB A 150 picosecond molecular dynamics computer simulation of the C-terminal fragment of the L7/L12 ribosomal protein from Escherichia coli is reported. The molecular dynamics results are compared with the available high-resolution X-ray data in terms of **atomic** positions, **distances** and positional fluctuations. Good agreement is found between the molecular dynamics results and the X-ray data. The form and parameters of the interaction potential **energy function** and the procedures for deriving it are discussed. Some current misunderstandings concerning the ways of evaluating the efficiency of molecular dynamics algorithms and of application of bond-length constraints in protein simulations are cleared up. The 150 picosecond trajectory has been scanned in a search for correlated motions within and between secondary structure elements. The beta-strands have diffusional stretching modes, and uncorrelated transversal displacements. The dynamic analysis of alpha-helices shows a variety of features. The atomic fluctuations differ between the helix ends; this effect reflects long time-scale motions. Two alpha-helices, alpha A and alpha C, show diffusive longitudinal stretching modes. The third helix, alpha B, has a correlated asymmetric longitudinal stretching; the N-terminal part dominates this behaviour. Furthermore, alpha B presents a librational motion with respect to the other parts of the molecule with a frequency of approximately 5 cm-1. This motion is coupled to helix stretching. Interestingly, the regions of highly conserved residues contain the most mobile parts of the molecule.

L166 ANSWER 10 OF 70 MEDLINE on STN
ACCESSION NUMBER: 2002409071 MEDLINE
DOCUMENT NUMBER: 22153201 PubMed ID: 12163065
TITLE: Energy estimation in protein design.
AUTHOR: Mendes Joaquim; Gueroïs Raphael; Serrano Luis
CORPORATE SOURCE: European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69117 Heidelberg, Germany.
SOURCE: CURRENT OPINION IN STRUCTURAL BIOLOGY, (2002 Aug) 12 (4) 441-6. Ref: 73
Journal code: 9107784. ISSN: 0959-440X.
PUB. COUNTRY: England: United Kingdom
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
General Review; (REVIEW)
(REVIEW, TUTORIAL)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 200301

ENTRY DATE: Entered STN: 20020807
Last Updated on STN: 20030130
Entered Medline: 20030129

AB The progress achieved by several groups in the field of computational protein design shows that successful design methods include two major features: efficient algorithms to deal with the combinatorial exploration of sequence space and optimal **energy functions** to rank sequences according to their fitness for the given fold.

L166 ANSWER 11 OF 70 MEDLINE on STN
ACCESSION NUMBER: 2001402918 MEDLINE
DOCUMENT NUMBER: 21347075 PubMed ID: 11455544
TITLE: Determination of conformational equilibrium of peptides in solution by NMR spectroscopy and theoretical conformational analysis: application to the calibration of **mean-field** solvation models.
AUTHOR: Groth M; Malicka J; Rodziewicz- Motowidlo S; Czaplewski C; Klaudel L; Wiczk W; Liwo A
CORPORATE SOURCE: Faculty of Chemistry University of Gdansk Sobieskiego 18 80-952 Gdansk Poland.
SOURCE: BIOPOLYMERS, (2001) 60 (2) 79-95.
Journal code: 0372525. ISSN: 0006-3525.
PUB. COUNTRY: United States
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 200109
ENTRY DATE: Entered STN: 20010924
Last Updated on STN: 20010924
Entered Medline: 20010920

AB Peptides occur in solution as ensembles of conformations rather than in a fixed conformation. The existing **energy functions** are usually inadequate to predict the conformational equilibrium in solution, because of failure to account properly for solvation, if the solvent is not considered explicitly (which is usually prohibitively expensive). NMR data are therefore widely incorporated into theoretical conformational analysis. Because of conformational flexibility, restrained molecular dynamics (with restraints derived from NMR data), which is usually applied to determine protein conformation is of limited use in the case of peptides. Instead, (a) the restraints are averaged within predefined time windows during molecular dynamics (MD) simulations (time averaging), (b) multiple-copy MD simulations are carried out and the restraints are averaged over the copies (ensemble averaging), or (c) a representative ensemble of sterically feasible conformations is generated and the weights of the conformations are then fitted so that the computed average observables match the experimental data (weight fitting). All these approaches are briefly discussed in this article. If an adequate force field is used, conformations with large statistical weights obtained from the weight-fitting procedure should also have low energies, which can be implemented in force field calibration. Such a procedure is particularly attractive regarding the parameterization of the solvation energy in nonaqueous solvents, e.g., dimethyl sulfoxide, for which thermodynamic solvation data are scarce. A method for calibration of solvation parameters in dimethyl sulfoxide, which is based on this principle was recently proposed by C. Baysal and H. Meirovitch (Journal of the American Chemical Society, 1998, Vol. 120, pp. 800--812), in which the energy gap between the conformations compatible with NMR data and the alternative conformations is maximized. In this work we propose an alternative method based on the principle that the best-fitting statistical weights of conformations should match the Boltzmann weights computed with the force field applied. Preliminary results obtained using three test peptides of varying conformational mobility: H-Ser(1)-Pro(2)-Lys(3)-Leu(4)-OH, Ac-Tyr(1)-D-Phe(2)-Ser(3)-Pro(4)-Lys(5)-

Leu(6)-NH(2), and cyclo(Tyr(1)-D-Phe(2)-Ser(3)-Pro(4)-Lys(5)-Leu(6)) are presented.

L166 ANSWER 12 OF 70 MEDLINE on STN
ACCESSION NUMBER: 2001087349 MEDLINE
DOCUMENT NUMBER: 20574027 PubMed ID: 11124040
TITLE: The SH3-fold family: experimental evidence and prediction of variations in the folding pathways.
AUTHOR: Guerois R; Serrano L
CORPORATE SOURCE: EMBL, Meyerhofstrasse 1, Heidelberg, 69117, Germany.. guerois@embl-heidelberg.de
SOURCE: JOURNAL OF MOLECULAR BIOLOGY, (2000 Dec 15) 304 (5) 967-82. Journal code: 2985088R. ISSN: 0022-2836.
PUB. COUNTRY: ENGLAND: United Kingdom
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 200101
ENTRY DATE: Entered STN: 20010322
Last Updated on STN: 20010322
Entered Medline: 20010118

AB To investigate the relationships between protein topology, amino acid sequence and folding mechanisms, the folding transition state of the Sso7d protein has been characterised both experimentally and theoretically. Although Sso7d protein has a similar topology to that of the SH3 domains, the structure of its transition state is different from that of alpha-spectrin and src SH3 domains previously studied. The folding algorithm, Fold-X, including an **energy function** with specific sequence features, accounts for these differences and reproduces with a good agreement the set of experimental phi(double dagger-U) values obtained for the three proteins. Our analysis shows that taking into account sequence features underlying protein topology is critical for an accurate prediction of the folding process.

L166 ANSWER 13 OF 70 MEDLINE on STN
ACCESSION NUMBER: 2000406868 MEDLINE
DOCUMENT NUMBER: 20274101 PubMed ID: 10813838
TITLE: A novel exhaustive search algorithm for predicting the conformation of polypeptide segments in proteins.
AUTHOR: Deane C M; Blundell T L
CORPORATE SOURCE: Department of Biochemistry, University of Cambridge, United Kingdom.
SOURCE: PROTEINS, (2000 Jul 1) 40 (1) 135-44. Journal code: 8700181. ISSN: 0887-3585.
PUB. COUNTRY: United States
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 200008
ENTRY DATE: Entered STN: 20000901
Last Updated on STN: 20000901
Entered Medline: 20000822

AB We present a fast ab initio method for the prediction of local conformations in proteins. The program, PETRA, selects polypeptide fragments from a computer-generated database (APD) encoding all possible peptide fragments up to twelve amino acids long. Each fragment is defined by a representative set of eight straight phi/psi pairs, obtained iteratively from a trial set by calculating how fragments generated from them represent the protein databank (PDB). Ninety-six percent (96%) of length five fragments in crystal structures, with a resolution better than 1.5 A and less than 25% identity, have a conformer in the database with

less than 1 A root-mean-square deviation (rmsd). In order to select segments from APD, PETRA uses a set of simple rule-based filters, thus reducing the number of potential conformations to a manageable total. This reduced set is scored and sorted using rmsd fit to the anchor regions and a knowledge-based **energy function** dependent on the sequence to be modelled. The best scoring fragments can then be optimized by minimization of contact potentials and rmsd fit to the core model. The quality of the prediction made by PETRA is evaluated by calculating both the differences in rmsd and backbone torsion angles between the final model and the native fragment. The average rmsd ranges from 1.4 A for three residue loops to 3.9 A for eight residue loops.

L166 ANSWER 14 OF 70 MEDLINE on STN
ACCESSION NUMBER: 2001049413 MEDLINE
DOCUMENT NUMBER: 20406621 PubMed ID: 10952089
TITLE: A novel computational method for predicting the transmembrane structure of G-protein coupled receptors: application to human C5aR and C3aR.
AUTHOR: Rayan A; Siew N; Cherno-Schwartz S; Matzner Y; Bautsch W; Goldblum A
CORPORATE SOURCE: The Department of Medicinal Chemistry, School of Pharmacy, Hebrew University of Jerusalem, Israel.
SOURCE: RECEPTORS AND CHANNELS, (2000) 7 (2) 121-37.
Journal code: 9315376. ISSN: 1060-6823.
PUB. COUNTRY: Switzerland
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 200012
ENTRY DATE: Entered STN: 20010322
Last Updated on STN: 20010322
Entered Medline: 20001214

AB A novel algorithm was applied to the sequences of bacteriorhodopsin (BRh), of rhodopsin (Rh), and of the two human anaphylatoxin receptors, C5a-receptor (hC5aR) and C3a-receptor (hC3aR), that predicts their transmembrane domains (TMD) according to energy criteria alone, on the basis of their sequences and a template structure for each. Two consecutive criteria were applied for the predictions: the first is hydrophobicity of a sequence of residues, which determines the candidate stretches of residues that form one of the transmembrane helices. The second criterion is an **energy function** composed of inter residue contact energies, of hydrophobic contributions due to membrane exposure and of the interactions of a few residues with the phospholipid head groups. The sequence of candidate residues for each helix is longer than that of the template, and is finally determined by threading each of the candidate stretches on each of the template helices and evaluating the energy for all possible configurations. Contact energies between residues were taken from a database (Miyazawa S and Jernigan RL (1996) J Mol Biol 256 623-44). The algorithm predicts well the TMD structure of BRh based on its own template, and the TMD structure of Rh conforms well with the model of Baldwin et al (Baldwin JM Schertler GFX and Unger VM (1997) J Biol Chem 272 144-64). Results for the construction of the TMD of hC5aR and hC3aR were compared, employing the template structure of Rh. Most of the results for these receptors are in accord with alignments and with mutation experiments on hC5aR and hC3aR. The predictions may serve as a basis for future mutagenesis experiments of these receptors.

L166 ANSWER 15 OF 70 MEDLINE on STN
ACCESSION NUMBER: 1999226799 MEDLINE
DOCUMENT NUMBER: 99226799 PubMed ID: 10211817
TITLE: Predicting the structures of 18 peptides using Geocore.

AUTHOR: Ishikawa K; Yue K; Dill K A
CORPORATE SOURCE: Central Research Laboratories, Ajinomoto Co., Kawasaki, Japan.
SOURCE: PROTEIN SCIENCE, (1999 Apr) 8 (4) 716-21.
Journal code: 9211750. ISSN: 0961-8368.
PUB. COUNTRY: United States
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 199907
ENTRY DATE: Entered STN: 19990727
Last Updated on STN: 20000303
Entered Medline: 19990714
AB We describe an extensive test of Geocore, an ab initio peptide folding algorithm. We studied 18 short molecules for which there are structures in the Protein Data Bank; chains are up to 31 monomers long. Except for the very shortest peptides, an extremely simple **energy function** is sufficient to discriminate the true native state from more than 10(8) lowest energy conformations that are searched explicitly for each peptide. A high incidence of native-like structures is found within the best few hundred conformations generated by Geocore for each amino acid sequence. Predictions improve when the number of discrete phi/psi choices is increased.

L166 ANSWER 16 OF 70 MEDLINE on STN
ACCESSION NUMBER: 1998399438 MEDLINE
DOCUMENT NUMBER: 98399438 PubMed ID: 9729739
TITLE: Computer search algorithms in protein modification and design.
AUTHOR: Desjarlais J R; Clarke N D
CORPORATE SOURCE: Department of Chemistry, Pennsylvania State University, University Park 16802, USA.. jrd@chem.psu.edu
SOURCE: CURRENT OPINION IN STRUCTURAL BIOLOGY, (1998 Aug) 8 (4) 471-5. Ref: 40
Journal code: 9107784. ISSN: 0959-440X.
PUB. COUNTRY: ENGLAND: United Kingdom
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
General Review; (REVIEW)
(REVIEW, TUTORIAL)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 199811
ENTRY DATE: Entered STN: 19990106
Last Updated on STN: 19990106
Entered Medline: 19981113
AB The computer-aided design of protein sequences requires efficient search algorithms to handle the enormous combinatorial complexity involved. A variety of different algorithms have now been applied with some success. The choice of algorithm can influence the representation of the problem in several important ways--the discreteness of the configuration, the types of energy terms that can be used and the ability to find the global minimum energy configuration. The use of dead end elimination to design the complete sequence for a small protein motif and the use of genetic and **mean-field** algorithms to design hydrophobic cores for proteins represent the major themes of the past year.

L166 ANSWER 17 OF 70 MEDLINE on STN
ACCESSION NUMBER: 1998051965 MEDLINE
DOCUMENT NUMBER: 98051965 PubMed ID: 9390292
TITLE: The native sequence determines sidechain packing in a protein, but does optimal sidechain packing determine the native sequence?.
AUTHOR: Koehl P; Delarue M

CORPORATE SOURCE: UPR 9003 du CNRS, Graffenstaden, France.
SOURCE: PACIFIC SYMPOSIUM ON BIOCOMPUTING, (1997) 198-209.
Journal code: 9711271.
PUB. COUNTRY: Singapore
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 199801
ENTRY DATE: Entered STN: 19980129
Last Updated on STN: 19980129
Entered Medline: 19980115

AB Globular proteins have highly compact structures and the corresponding packing interactions are widely considered as the principal determinant of the native structure. It is therefore important that theoretical approaches to protein design explicitly take in account packing, which requires that a full atomic representation of the designed protein is maintained. As a first step towards this goal, we have developed in this report an inverse folding algorithm with the aim of specifically designing amino acid sequences which optimise sidechain packing for a given protein fold. The design is performed by a global Monte Carlo optimisation in sequence space, with constant amino acid composition and a full-atom representation of the various protein models. Packing is defined by a Lennard-Jones potential. The program was tested by designing stable sequence variants for the chymotrypsin inhibitor fold. The final protein models showed an increase in intramolecular atomic contacts and a decrease in the overall volume compared to the native structure. Starting from the backbone only of the target structure, the algorithm did gradually retrieve reliable though limited sequence information. Higher compatibility might be achieved by improving the potential, however our results suggest that packing interactions are an essential element of a yet-to-be-defined successful **energy function** for protein design.

L166 ANSWER 18 OF 70 MEDLINE on STN
ACCESSION NUMBER: 96291583 MEDLINE
DOCUMENT NUMBER: 96291583 PubMed ID: 8728655
TITLE: **Mean-field** minimization methods for biological macromolecules.
AUTHOR: Koehl P; Delarue M
CORPORATE SOURCE: UPR 9003 du CNRS, Ecole Superieure de Biotechnologie de Strasbourg, France.. koehl@bali.u-strasbg.fr
SOURCE: CURRENT OPINION IN STRUCTURAL BIOLOGY, (1996 Apr) 6 (2) 222-6. Ref: 36
Journal code: 9107784. ISSN: 0959-440X.
PUB. COUNTRY: ENGLAND: United Kingdom
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
General Review; (REVIEW)
(REVIEW, TUTORIAL)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 199609
ENTRY DATE: Entered STN: 19961008
Last Updated on STN: 19961008
Entered Medline: 19960924

AB Simulations of macromolecular structures involve the minimization of a potential-**energy function** that presents many local minima. **Mean-field** theory provides a tool that enables us to escape these minima, by enhancing sampling in conformational space. The number of applications of this technique has increased significantly over the past year, enabling problems with protein-homology modelling and inverted protein structure prediction to be solved.

L166 ANSWER 19 OF 70 MEDLINE on STN

ACCESSION NUMBER: 95018272 MEDLINE
DOCUMENT NUMBER: 95018272 PubMed ID: 7932758
TITLE: A graph-theoretic approach to the identification of three-dimensional patterns of amino acid side-chains in protein structures.
AUTHOR: Artymiuk P J; Poirrette A R; Grindley H M; Rice D W; Willett P
CORPORATE SOURCE: Krebs Institute for Biomolecular Research Department of Molecular Biology and Biotechnology, University of Sheffield, Western Bank, U.K.
SOURCE: JOURNAL OF MOLECULAR BIOLOGY, (1994 Oct 21) 243 (2) 327-44. Journal code: 2985088R. ISSN: 0022-2836.
PUB. COUNTRY: ENGLAND: United Kingdom
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 199411
ENTRY DATE: Entered STN: 19941222
Last Updated on STN: 20000303
Entered Medline: 19941110

AB This paper discusses the use of graph-theoretic methods for the representation and searching of three-dimensional patterns of side-chains in protein structures. The position of a side-chain is represented by pseudo-atoms, and the relative positions of pairs of side-chains by the distances between them. This description of the geometry can be represented by a labelled graph in which the nodes and the edges of the graph represent the pseudo-atoms and the sets of inter-pseudo-**atomic distances**, respectively. Given such a representation, a protein can be searched for the presence of a user-defined query pattern of side-chains by means of a subgraph-isomorphism algorithm which is implemented in the program ASSAM. Experiments with one such algorithm, that due to Ullmann, show that it provides both an effective and a highly efficient way of searching for patterns of side-chains. The method is illustrated by searches for the serine protease catalytic triad, for residues involved in the catalytic activity of staphyloccocal nuclease, and for the zinc-binding side-chains of thermolysin. The catalytic triad pattern search revealed the existence of a second Asp-His-Ser triad-like arrangement of residues in trypsinogen and chymotrypsinogen, in addition to the catalytic residues. In addition the program can be used to search for hypothetical patterns, as is shown for a pattern of three tryptophan side-chains. These searches demonstrate that the search algorithm can successfully retrieve the great majority of the expected proteins, as well as other, previously unreported proteins that contain the pattern of interest.

L166 ANSWER 20 OF 70 PASCAL on STN
ACCESSION NUMBER: 1992-0583528 PASCAL
TITLE (IN ENGLISH): Conjugate peak refinenement : an **algorithm** for finding reaction paths and accurate transition states in systems with many degrees of freedom
AUTHOR: FISCHER S.; KARPLUS M.
CORPORATE SOURCE: Harvard univ., dep. chemistry, comitte higher degrees biophysics, Cambridge MA 01238, United States
SOURCE: Chemical physics letters, (1992), 194(3), 252-261, 22 refs.
ISSN: 0009-2614 CODEN: CHPLBC
DOCUMENT TYPE: Journal
BIBLIOGRAPHIC LEVEL: Analytic
COUNTRY: Netherlands
LANGUAGE: English
AVAILABILITY: INIST-13494, 354000028608320200

AB An **algorithm** is presented for determining multi-dimensional

reaction **coordinates** between two known conformers. Only the **energy function** and its gradient are required. The resulting paths follow the adiabatic energy valleys and have energy maxima that are true saddle points, which can be multiple along each path. The method is suitable for the study of complex isomerization reactions, including allosteric transitions in **proteins** and more general **conformational** changes of macromolecules

L166 ANSWER 21 OF 70 PASCAL on STN

ACCESSION NUMBER: 1992-0020748 PASCAL
TITLE (IN ENGLISH): Constrained optimization and **protein structure** determination
AUTHOR: EISENFELD J.; VAJDA S.; SUGAR I.; DELISI C.
CORPORATE SOURCE: Mount Sinai school medicine, dep. biomathematical sci., New York NY 10029, United States
SOURCE: American journal of physiology. Cell physiology, (1991), 30(2), C376-C386, 30 refs.
ISSN: 0363-6143 CODEN: AJPCDD
DOCUMENT TYPE: Journal
BIBLIOGRAPHIC LEVEL: Analytic
COUNTRY: United States
LANGUAGE: English
SUMMARY LANGUAGE: English
AVAILABILITY: INIST-670 B, 354000012800570220

AB Energy minimization is one of the main approaches to the **computational** determination of macromolecular structure. Due to the approximations in the empirical free-**energy functions** and due to the **computational** difficulties in locating their global minima, the problem is at present intractable when the only information available is the sequence of subunits forming the molecule. A less-demanding problem in terms of both physics and mathematics is constrained optimization, which uses additional but incomplete experimental information such as **distances** between certain **atoms**. This paper reviews methods for generating molecular structure using bond lengths and angles as variables and shows how the structure can be fully specified in termsof local geometry

L166 ANSWER 22 OF 70 BIOTECHNO on STN DUPLICATE

ACCESSION NUMBER: 2003:36505431 BIOTECHNO
TITLE: Cyclic **coordinate** descent: A robotics **algorithm** for protein loop closure
AUTHOR: Canutescu A.A.; Dunbrack Jr. R.L.
CORPORATE SOURCE: Dr. R.L. Dunbrack Jr., Institute for Cancer Research, Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, PA 19111, United States.
E-mail: RL_Dunbrack@fccc.edu
SOURCE: Protein Science, (01 MAY 2003), 12/5 (963-972), 40 reference(s)
CODEN: PRCIEI ISSN: 0961-8368
DOCUMENT TYPE: Journal; Article
COUNTRY: United States
LANGUAGE: English
SUMMARY LANGUAGE: English

AB In **protein structure** prediction, it is often the case that a protein segment must be adjusted to connect two fixed segments. This occurs during loop structure prediction in homology modeling as well as in ab initio structure prediction. Several **algorithms** for this purpose are based on the inverse Jacobian of the distance constraints with respect to dihedral angle degrees of freedom. These **algorithms** are sometimes unstable and fail to converge. We present an **algorithm** developed originally for inverse

kinematics applications in robotics. In robotics, an end effector in the form of a robot hand must reach for an object in space by altering adjustable joint angles and arm lengths. In loop prediction, dihedral angles must be adjusted to move the C-terminal residue of a segment to superimpose on a fixed anchor residue in the **protein structure**. The **algorithm**, referred to as cyclic **coordinate** descent or CCD, involves adjusting one dihedral angle at a time to minimize the sum of the squared **distances** between three backbone **atoms** of the moving C-terminal anchor and the corresponding atoms in the fixed C-terminal anchor. The result is an equation in one variable for the proposed change in each dihedral. The **algorithm** proceeds iteratively through all of the adjustable dihedral angles from the N-terminal to the C-terminal end of the loop. CCD is suitable as a component of loop prediction methods that generate large numbers of trial structures. It succeeds in closing loops in a large test set 99.79% of the time, and fails occasionally only for short, highly extended loops. It is very fast, closing loops of length 8 in 0.037 sec on average.

L166 ANSWER 23 OF 70 BIOTECHNO COPYRIGHT 2003 Elsevier Science B.V. on STN DUPLICATE

ACCESSION NUMBER: 2001:33027663 BIOTECHNO
TITLE: SuperStar: Improved knowledge-based interaction fields for protein binding sites
AUTHOR: Verdonk M.L.; Cole J.C.; Watson P.; Gillet V.; Willett P.
CORPORATE SOURCE: M.L. Verdonk, Astex Technology Ltd., 250 Cambridge Science Park, Cambridge CB4 0WE, United Kingdom. E-mail: m.verdonk@astex-technology.com
SOURCE: Journal of Molecular Biology, (30 MAR 2001), 307/3 (841-859), 51 reference(s)
CODEN: JMOBAK ISSN: 0022-2836
DOCUMENT TYPE: Journal; Article
COUNTRY: United Kingdom
LANGUAGE: English
SUMMARY LANGUAGE: English

AB SuperStar is an empirical method for identifying interaction sites in proteins, based entirely on experimental information about non-bonded interactions occurring in small-molecule crystal structures, taken from the IsoStar database. We describe recent modifications and additions to SuperStar, validating the results on a test set of 122 X-ray **structures** of **protein**-ligand complexes. In this validation, propensity maps are generated for all the binding sites of these proteins, using four different probes: a charged NH.sub.3.sup.+ nitrogen atom, a carbonyl oxygen atom, a hydroxyl oxygen atom and a methyl carbon atom. Next, the maps are compared with the experimentally observed positions of ligand atoms of these types. A peak-searching **algorithm** is introduced that highlights potential interaction hot spots. For the three hydrogen-bonding probes - NH.sub.3.sup.+ nitrogen atom, carbonyl oxygen atom and hydroxyl oxygen **atom** - the average **distance** from the ligand **atom** to the nearest SuperStar peak is 1.0-1.2 .ANG. (0.8-1.0 .ANG. for solvent-inaccessible ligand atoms). For the methyl carbon **atom** probe, this **distance** is about 1.5 .ANG., probably because interactions to methyl groups are much less directional. The most important addition to SuperStar is the enabling of propensity maps around metal centres - Ca.sup.2.sup.+, Mg.sup.2.sup.+ and Zn.sup.2.sup.+ - in protein binding sites. The results are validated on a test set of 24 protein-ligand complexes that have a metal ion in their binding site. Coordination geometries are derived automatically, using only the protein atoms that **coordinate** to the metal ion. The correct coordination geometry is derived in approximately 75% of the cases. If the derived geometry is assumed during the SuperStar calculation, the average **distance**

from a ligand **atom** coordinating to the metal ion to the nearest
peak in the propensity map for an oxygen probe is 0.87(7) .ANG.. If the
correct coordination geometry is imposed, this distance reduces to
0.59(7) .ANG.. This indicates that the SuperStar predictions around
metal-binding sites are at least as good as those around other protein
groups. Using clustering techniques, a non-redundant set of probes is
selected from the set of probes available in the IsoStar database. The
performance in SuperStar of all these probes is tested on the test set of
protein-ligand complexes. With the exception of the "ether oxygen" probe
and the "any NH.sup.+" probe, all new probes perform as well as the four
probes introduced first. .COPYRGT. 2001 Academic Press.

L166 ANSWER 24 OF 70 BIOTECHNO COPYRIGHT 2003 Elsevier Science B.V. on STN
DUPLICATE

ACCESSION NUMBER: 1999:29128152 BIOTECHNO
TITLE: **Protein** tertiary **structure** prediction using a branch and bound **algorithm**
AUTHOR: Eyrich V.A.; Standley D.M.; Felts A.K.; Friesner R.A.
CORPORATE SOURCE: R.A. Friesner, Department of Chemistry, Center for Biomolecular Simulation, Columbia University, New York, NY 10027, United States.
E-mail: rich@chm.columbia.edu
SOURCE: Proteins: Structure, Function and Genetics, (01 APR 1999), 35/1 (41-57), 34 reference(s)
CODEN: PSFGEY ISSN: 0887-3585
DOCUMENT TYPE: Journal; Article
COUNTRY: United States
LANGUAGE: English
SUMMARY LANGUAGE: English

AB We report a new method for predicting **protein** tertiary
structure from sequence and secondary structure information. The
predictions result from global optimization of a potential **energy**
function, including van der Waals, hydrophobic, and excluded
volume terms. The optimization **algorithm**, which is based on the
.alpha.BB method developed by Floudas and coworkers (Costas and Floudas,
J Chem Phys 1994;100:1247-1261), uses a reduced model of the protein and
is implemented in both distance and dihedral angle space, enabling a
side-by-side comparison of methodologies. For a set of eight small
proteins, representing the three basic typesall .alpha., all .beta., and
mixed .alpha./.beta.--the **algorithm** locates low-energy
native-like structures (less than 6.ANG. root mean square deviation from
the native **coordinates**) starting from an unfolded state. Serial
and parallel implementations of this methodology are discussed.

L166 ANSWER 25 OF 70 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN

ACCESSION NUMBER: 2003:242466 BIOSIS
DOCUMENT NUMBER: PREV200300242466
TITLE: Addition of side chains to a known backbone with defined side-chain centroids.
AUTHOR(S): Kazmierkiewicz, Rajmund; Liwo, Adam; Scheraga, Harold A. (1)
CORPORATE SOURCE: (1) Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, NY, 14853-1301, USA: has5@cornell.edu USA
SOURCE: Biophysical Chemistry, (2003) Vol. 100, No. 1-3, pp. 261-280. print.
ISSN: 0301-4622.
DOCUMENT TYPE: Article
LANGUAGE: English

AB An automatic procedure is proposed for adding side chains to a protein
backbone; it is based on optimization of a simplified **energy**
function for peptide side chains, given its backbone and positions
of side-chain centroids. The energy is expressed as a sum of the energies

of interaction between side chains, and a harmonic penalty function accounting for the preservation of the positions of the Calpha atoms and the side-chain centroids. The energy of side-chain interactions is calculated with the soft-sphere ECEPP/3 potential. A Monte Carlo search is carried out to explore all possible side-chain orientations within a fixed backbone and side-chain centroid positions. The initial, usually extended, side-chain conformations are taken directly from the ECEPP/3 database. The procedure was tested on six experimental (X-ray or NMR) **structures** : immunoglobulin binding **protein** (PDB code 1IGD, an alpha+beta-protein); transcription factor PML (PDB code 1BOR, a 49-104 fragment of the ring finger domain, predominantly beta-protein); bovine pancreatic trypsin inhibitor (crystal form II) (PDB code 1BPI, an alpha+beta-protein); the monomer of human deoxyhemoglobin (PDB code 1BZ0, an alpha-helical structure); chain A of alcohol dehydrogenase from Drosophila lebanonensis (PDB code 1A4U); as well as on the 10-55 portion of the B domain of staphylococcal protein A (PDB code 1BDD). In all cases except 1BPI, the data for the **algorithm** (i.e. the backbone or Calpha **coordinates** and the positions of side-chain centroids) were taken from the experimental **structures**. For **protein** A, the Calpha **coordinates** and positions of side-chain centroids were also taken from the 1.9-ANG-resolution model predicted by the UNRES force field. In all comparisons with experimental structures, complete side-chain geometry was reconstructed with a root-mean-square (RMS) deviation of approximately 0.6-0.9 ANG from the heavy atoms when complete backbone and side-chain-centroid **coordinates** were used in reconstruction, or approximately 1.0 ANG when the Calpha and centroid **coordinates** were used.

ACCESSION NUMBER: 2003:102888 BIOSIS
DOCUMENT NUMBER: PREV200300102888
TITLE: Beyond the rotamer library: Genetic **algorithm** combined with the disturbing mutation process for upbuilding protein side-chains.
AUTHOR(S): Liu, Zhijie; Jiang, Lin; Gao, Ying; Liang, Shide; Chen, Hao; Han, Yuzhen; Lai, Luhua (1)
CORPORATE SOURCE: (1) College of Chemistry and Molecular Engineering, and State Key Laboratory for Structural Chemistry of Unstable and Stable Species, Institute of Physical Chemistry, Peking University, Beijing, 100871, China: lhlai@pku.edu.cn, lai@mdl.ipc.pku.edu.cn China
SOURCE: Proteins, (January 1 2003) Vol. 50, No. 1, pp. 49-62. print.
ISSN: 0887-3585.
DOCUMENT TYPE: Article
LANGUAGE: English
AB The disturbing genetic **algorithm**, incorporating the disturbing mutation process into the genetic **algorithm** flow, has been developed to extend the searching space of side-chain conformations and to improve the quality of the rotamer library. Moreover, the growing generation amount idea, simulating the real situation of the natural evolution, is introduced to improve the searching speed. In the calculations using the pseudo energy scoring function of the root mean squared deviation, the disturbing genetic **algorithm** method has been shown to be highly efficient. With the real **energy function** based on AMBER force field, the program has been applied to rebuilding side-chain conformations of 25 high-quality crystallographic **structures** of single-**protein** and **protein-protein** complexes. The averaged root mean standard deviation of atom **coordinates** in side-chains and veracities of the torsion angles of chi1 and chi1+chi2 are 1.165 ANG, 88.2 and 72.9% for the buried residues, respectively, and 1.493 ANG, 79.2 and 64.7% for all residues, showing that the method has equal precision to the program SCWRL, whereas

it performs better in the prediction of buried residues and protein-protein interfaces. This method has been successfully used in redesigning the interface of the Basnase-Barstar complex, indicating that it will have extensive application in **protein** design, **protein** sequence and **structure** relationship studies, and research on **protein**-protein interaction.

L166 ANSWER 27 OF 70 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
ACCESSION NUMBER: 2002:180203 BIOSIS
DOCUMENT NUMBER: PREV200200180203
TITLE: Soft protein-protein docking in internal **coordinates**.
AUTHOR(S): Fernandez-Recio, Juan; Totrov, Maxim; Abagyan, Ruben (1)
CORPORATE SOURCE: (1) Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, TPC-28, La Jolla, CA, 92037: abagyan@scripps.edu USA
SOURCE: Protein Science, (February, 2002) Vol. 11, No. 2, pp. 280-291. print.
ISSN: 0961-8368.
DOCUMENT TYPE: Article
LANGUAGE: English
AB The association of two biological macromolecules is a fundamental biological phenomenon and an unsolved theoretical problem. Docking methods for ab initio prediction of association of two independently determined **protein structures** usually fail when they are applied to a large set of complexes, mostly because of inaccuracies in the scoring function and/or difficulties on simulating the rearrangement of the interface residues on binding. In this work we present an efficient pseudo-Brownian rigid-body docking procedure followed by Biased Probability Monte Carlo Minimization of the ligand interacting side-chains. The use of a soft interaction **energy function** precalculated on a grid, instead of the explicit energy, drastically increased the speed of the procedure. The method was tested on a benchmark of 24 protein-protein complexes in which the three-dimensional structures of their subunits (bound and free) were available. The rank of the near-native conformation in a list of candidate docking solutions was <20 in 85% of complexes with no major backbone motion on binding. Among them, as many as 7 out of 11 (64%) protease-inhibitor complexes can be successfully predicted as the highest rank conformations. The presented method can be further refined to include the binding site predictions and applied to the structures generated by the structural proteomics projects. All scripts are available on the Web.

L166 ANSWER 28 OF 70 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
ACCESSION NUMBER: 1994:342020 BIOSIS
DOCUMENT NUMBER: PREV199497355020
TITLE: Measuring residue associations in **protein structures**: Possible implications for **protein** folding.
AUTHOR(S): Karlin, Samuel (1); Zuker, Michael; Brocchieri, Luciano (1)
CORPORATE SOURCE: (1) Dep. Math., Stanford Univ., Stanford, CA 94305-2125 USA
SOURCE: Journal of Molecular Biology, (1994) Vol. 239, No. 2, pp. 227-248.
ISSN: 0022-2836.
DOCUMENT TYPE: Article
LANGUAGE: English
AB We propose a number of distance measures between residues in **protein structures** based on average, minimum and maximum **distances** of all **atom** (backbone and side-chain) **coordinates** or with respect to side-chain atom **coordinates** only. The d-1-distance (D-1-distance) refers to the average distance between side-chain (backbone and side-chain) atoms of a residue pair in a given structure. The d-m-distance (D-m-distance) refers to the minimum

distance between side-chain **atoms** (non trivial minimum **distance** between all **atoms** of a residue pair). For each distance measure, averaging and normalizing over representative **protein structures**, association values and closeness orderings for all amino acid types are determined. The expected associations of side-chain interactions between oppositely charged residues, among hydrophobic residues and of cysteine with cysteine are confirmed. Several surprising associations are observed relative to (1) the aromatic residues tyrosine and tryptophan, but not phenylalanine (2) multiple histidine residues; (3) asymmetries of arginine versus lysine, aspartate versus glutamate, alanine versus glycine, and asparagine versus glutamine; (4) absence of correlations of alpha-carbon distances with side-chain **distances**. The all **atoms** D-1-**distance** attractions are dominated by steric relationships, with glycine and alanine significantly close to all amino acids, whereas large residues are under-associated with all residue types. In contrast, for the closeness ordering corresponding to the minimum side-chain d-m-distance, glycine arid alanine are among the least associated. However, in the d-1-distance alanine is significantly close to all hydrophobic residues with the exception of tryptophan. The d-m-distance preferences display a pervasive attraction for tyrosine by almost all residue types, the prominence of tyrosine and tryptophan in cation-aromatic interactions, and the versatility of histidine in functionality. The principal findings suggest a new perspective on the early and intermediate stages of protein folding.

L166 ANSWER 29 OF 70 SCISEARCH COPYRIGHT 2003 THOMSON ISI on STN DUPLICATE 3

ACCESSION NUMBER:	2001:631669 SCISEARCH
THE GENUINE ARTICLE:	457DH
TITLE:	A **distance**-dependent **atomic** knowledge-based potential for improved **protein structure** selection
AUTHOR:	Lu H; Skolnick J (Reprint)
CORPORATE SOURCE:	Donald Danforth Plant Sci Ctr, Lab Computat Genomics, St Louis, MO USA (Reprint)
COUNTRY OF AUTHOR:	USA
SOURCE:	PROTEINS-STRUCTURE FUNCTION AND GENETICS, (15 AUG 2001) Vol. 44, No. 3, pp. 223-232. Publisher: WILEY-LISS, DIV JOHN WILEY & SONS INC, 605 THIRD AVE, NEW YORK, NY 10158-0012 USA. ISSN: 0887-3585.
DOCUMENT TYPE:	Article; Journal
LANGUAGE:	English
REFERENCE COUNT:	40

ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB A heavy **atom distance**-dependent knowledge-based pairwise potential has been developed. This statistical potential is first evaluated and optimized with the native structure z-scores from gapless threading. The potential is then used to recognize the native and near-native structures from both published decoy test sets, as well as decoys obtained from our group's **protein structure** prediction program. In the gapless threading test, there is an average z-score improvement of 4 units in the optimized atomic potential over the residue-based quasichemical potential. Examination of the z-scores for individual pairwise distance shells indicates that the specificity for the native **protein structure** is greatest at pairwise distances of 3.5-6.5 Angstrom, i.e., in the first solvation shell. On applying the current atomic potential to test sets obtained from the web, composed of native **protein** and decoy **structures,** the current generation of the potential performs better than residue-based potentials as well as the other published atomic potentials in the task of selecting native and near-native structures. This newly developed potential is also applied to structures of varying quality generated by

our group's **protein structure** prediction program. The current atomic potential tends to pick lower RMSD structures than do residue-based contact potentials. In particular, this atomic pairwise interaction potential has better selectivity especially for near-native structures. As such, it can be used to select near-native folds generated by structure prediction **algorithms** as well as for **protein structure** refinement. (C) 2001 Wiley-Liss, Inc.

L166 ANSWER 30 OF 70 SCISEARCH COPYRIGHT 2003 THOMSON ISI on STN DUPLICATE 9
ACCESSION NUMBER: 1998:168541 SCISEARCH
THE GENUINE ARTICLE: YY045
TITLE: An all-**atom distance**-dependent conditional probability discriminatory function for **protein structure** prediction
AUTHOR: Samudrala R; Moult J (Reprint)
CORPORATE SOURCE: UNIV MARYLAND, MARYLAND BIOTECHNOL INST, CTR ADV RES BIOTECHNOL, 9600 GUDELSKY DR, ROCKVILLE, MD 20850 (Reprint); UNIV MARYLAND, MARYLAND BIOTECHNOL INST, CTR ADV RES BIOTECHNOL, ROCKVILLE, MD 20850; UNIV MARYLAND, PROGRAM MOL & CELL BIOL, COLLEGE PK, MD 20742
COUNTRY OF AUTHOR: USA
SOURCE: JOURNAL OF MOLECULAR BIOLOGY, (6 FEB 1998) Vol. 275, No. 5, pp. 895-916.
Publisher: ACADEMIC PRESS LTD, 24-28 OVAL RD, LONDON, ENGLAND NW1 7DX.
ISSN: 0022-2836.
DOCUMENT TYPE: Article; Journal
FILE SEGMENT: LIFE
LANGUAGE: English
REFERENCE COUNT: 61
ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB We present a formalism to **compute** the probability of an amino acid sequence conformation being native-like, given a set of pairwise **atom-atom distances**. The formalism is used to derive three discriminatory functions with different types of representations for the atom-atom contacts observed in a database of **protein structures**. These functions include two virtual atom representations and one all-heavy atom representation. When applied to six different decoy sets containing a range of correct and incorrect conformations of amino acid sequences, the all-**atom distance**-dependent discriminatory function is able to identify correct from incorrect more often than the discriminatory functions using approximate representations. We illustrate the importance of using a detailed atomic description for obtaining the most accurate discrimination, and the necessity for testing discriminatory functions against a wide variety of decoys. The discriminatory function is also shown to be capable of capturing the fine details of atom-atom preferences. These results suggest that the all-**atom distance**-dependent discriminatory function will be useful for **protein structure** prediction and model refinement. (C) 1998 Academic Press Limited.

L166 ANSWER 31 OF 70 SCISEARCH COPYRIGHT 2003 THOMSON ISI on STN
ACCESSION NUMBER: 2003:313736 SCISEARCH
THE GENUINE ARTICLE: 665QV
TITLE: A geometric build-up **algorithm** for solving the molecular distance geometry problem with sparse distance data
AUTHOR: Dong Q F (Reprint); Wu Z J
CORPORATE SOURCE: Iowa State Univ, Dept Zool & Genet, Ames, IA 50010 USA (Reprint); Iowa State Univ, Dept Math, Ames, IA 50010 USA; Iowa State Univ, Grad Program Bioinformat & Computat Biol, Ames, IA 50010 USA

COUNTRY OF AUTHOR: USA
SOURCE: JOURNAL OF GLOBAL OPTIMIZATION, (JUL 2003) Vol. 26, No. 3, pp. 321-333.
Publisher: KLUWER ACADEMIC PUBL, VAN GODEWIJCKSTRAAT 30, 3311 GZ DORDRECHT, NETHERLANDS.
ISSN: 0925-5001.
DOCUMENT TYPE: Article; Journal
LANGUAGE: English
REFERENCE COUNT: 17
ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB Nuclear magnetic resonance (NMR) structure modeling usually produces a sparse set of inter-**atomic distances** in protein. In order to calculate the three-dimensional **structure** of **protein**, current approaches need to estimate all other "missing" distances to build a full set of distances. However, the estimation step is costly and prone to introducing errors. In this report, we describe a geometric build-up **algorithm** for solving **protein structure** by using only a sparse set of inter-**atomic distances**. Such a sparse set of distances can be obtained by combining NMR data with our knowledge on certain bond lengths and bond angles. It can also include confident estimations on some "missing" distances. Our **algorithm** utilizes a simple geometric relationship between **coordinates** and distances. The **coordinates** for each atom are calculated by using the **coordinates** of previously determined **atoms** and their **distances**. We have implemented the **algorithm** and tested it on several proteins. Our results showed that our **algorithm** successfully determined the **protein structures** with sparse sets of distances. Therefore, our **algorithm** reduces the need of estimating the "missing" distances and promises a more efficient approach to NMR structure modeling.

L166 ANSWER 32 OF 70 SCISEARCH COPYRIGHT 2003 THOMSON ISI on STN
ACCESSION NUMBER: 1998:721444 SCISEARCH
THE GENUINE ARTICLE: 119WL
TITLE: **Algorithmic** determination of core positions in the V-L and V-H domains of immunoglobulin molecules
AUTHOR: Gelfand I (Reprint); Kister A; Kulikowski C; Stoyanov O
CORPORATE SOURCE: RUTGERS STATE UNIV, DEPT MATH, NEW BRUNSWICK, NJ 08903 (Reprint); RUTGERS STATE UNIV, DEPT COMP SCI, NEW BRUNSWICK, NJ 08903
COUNTRY OF AUTHOR: USA
SOURCE: JOURNAL OF COMPUTATIONAL BIOLOGY, (FAL 1998) Vol. 5, No. 3, pp. 467-477.
Publisher: MARY ANN LIEBERT INC PUBL, 2 MADISON AVENUE, LARCHMONT, NY 10538.
ISSN: 1066-5277.
DOCUMENT TYPE: Article; Journal
FILE SEGMENT: LIFE
LANGUAGE: English
REFERENCE COUNT: 18
ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB We introduce a new **algorithmic** method for identifying the geometrical core of proteins that does not require the usual superposition of structures. A geometrical core is defined as the set of residues such that the C-alpha(I) - C-alpha(J) **atom distances** are identical in all **structures** of the **protein** family under study, where I and J are secondary structure positions in the structural units-strands, loops, or parts of them. The result of applying the **algorithm** to 53 Ig structures leads to the identification of two geometrical core sets of C-alpha atom positions for the V-L and V-H domains. Applications of the core sets are described.

L166 ANSWER 33 OF 70 SCISEARCH COPYRIGHT 2003 THOMSON ISI on STN
ACCESSION NUMBER: 1998:780730 SCISEARCH
THE GENUINE ARTICLE: 125VE
TITLE: Accuracy of side-chain prediction upon near-native protein backbones generated by ab initio folding methods
AUTHOR: Huang E S; Koehl P; Levitt M; Pappu R V; Ponder J W (Reprint)
CORPORATE SOURCE: WASHINGTON UNIV, SCH MED, DEPT BIOCHEM & MOL BIOPHYS, ST LOUIS, MO 63110 (Reprint); WASHINGTON UNIV, SCH MED, DEPT BIOCHEM & MOL BIOPHYS, ST LOUIS, MO 63110; STANFORD UNIV, SCH MED, DEPT BIOL STRUCT, STANFORD, CA 94305
COUNTRY OF AUTHOR: USA
SOURCE: PROTEINS-STRUCTURE FUNCTION AND GENETICS, (1 NOV 1998) Vol. 33, No. 2, pp. 204-217.
Publisher: WILEY-LISS, DIV JOHN WILEY & SONS INC, 605 THIRD AVE, NEW YORK, NY 10158-0012.
ISSN: 0887-3585.
DOCUMENT TYPE: Article; Journal
FILE SEGMENT: LIFE
LANGUAGE: English
REFERENCE COUNT: 53
ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB The ab initio folding problem can be divided into two sequential tasks of approximately equal **computational** complexity: the generation of native-like backbone folds and the positioning of side chains upon these backbones. The prediction of side-chain conformation in this context is challenging, because at best only the near-native global fold of the protein is known. To test the effect of displacements in the protein backbones on side-chain prediction for folds generated ab initio, sets of near-native backbones (less than or equal to 4 Angstrom C alpha RMS error) for four small proteins were generated by two methods. The steric environment surrounding each residue was probed by placing the side chains in the native conformation on each of these decoys, followed by torsion-space optimization to remove steric clashes on a rigid backbone, We observe that on average 40% of the chi 1 angles were displaced by 40 degrees or more, effectively setting the limits in accuracy for sidechain modeling under these conditions. Three different **algorithms** were subsequently used for prediction of side-chain conformation. The average prediction accuracy for the three methods was remarkably similar: 49% to 51% of the chi 1 angles were predicted correctly overall (33% to 36% of the chi 1+2 angles). Interestingly, when the inter-side-chain interactions were disregarded, the mean accuracy increased. A consensus approach is described, in which side-chain conformations are defined based on the most frequently predicted chi angles for a given method upon each set of near-native backbones, We find that consensus modeling, which de facto includes backbone flexibility, improves side-chain prediction: chi 1 accuracy improved to 51-54% (36-42% of chi 1+2). Implications of a consensus method for ab initio **protein structure** prediction are discussed. (C) 1998 Wiley-Liss, Inc.

L166 ANSWER 34 OF 70 SCISEARCH COPYRIGHT 2003 THOMSON ISI on STN
ACCESSION NUMBER: 97:511143 SCISEARCH
THE GENUINE ARTICLE: XH318
TITLE: Distance geometry based comparative modelling
AUTHOR: Aszodi A; Munro R E J; Taylor W R (Reprint)
CORPORATE SOURCE: NATL INST MED RES, DIV MATH BIOL, MILL HILL, LONDON NW7 1AA, ENGLAND (Reprint); NATL INST MED RES, DIV MATH BIOL, LONDON NW7 1AA, ENGLAND
COUNTRY OF AUTHOR: ENGLAND
SOURCE: FOLDING & DESIGN, (JUL 1997) Vol. 2, No. 3, pp. S3-S6.
Publisher: CURRENT BIOLOGY LTD, 34-42 CLEVELAND STREET, LONDON, ENGLAND W1P 6LB.
ISSN: 1359-0278.

DOCUMENT TYPE: Article; Journal
FILE SEGMENT: LIFE
LANGUAGE: English
REFERENCE COUNT: 24
ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB A distance geometry based protein modelling **algorithm** is presented which relies on the projection of simple model chain **coordinates** into Euclidean spaces with gradually decreasing dimensionality. Fast embedding was achieved by performing separate distance matrix projections on subsets of the model points. Structural equivalences between the unknown target and related **proteins** with known **structures** were deduced either from a mixed sequence/structure multiple alignment or from the output of various fold recognition (threading) approaches. These equivalences were mapped onto the model as structure-specific conserved C-alpha **atom distances** and secondary structure assignments. Additional nonspecific distance restraints derived from general stereochemical properties of folded protein chains were used to guide the modelling process. The method quickly constructed a large number of low-resolution models which could then serve as starting conformations for full-atom refinement. Structure predictions for some targets in the 'Asilomar Challenge' (GASPS) are presented to illustrate potential applications of the approach.

L166 ANSWER 35 OF 70 SCISEARCH COPYRIGHT 2003 THOMSON ISI on STN
ACCESSION NUMBER: 92:36356 SCISEARCH
THE GENUINE ARTICLE: GY176
TITLE: MSEED - A PROGRAM FOR THE RAPID ANALYTICAL DETERMINATION OF ACCESSIBLE SURFACE-AREAS AND THEIR DERIVATIVES
AUTHOR: PERROT G; CHENG B; GIBSON K D; VILA J; PALMER K A; NAYEEM A; MAIGRET B; SCHERAGA H A (Reprint)
CORPORATE SOURCE: CORNELL UNIV, BAKER LAB CHEM, ITHACA, NY, 14853; UNIV STRASBOURG 1, INST LEBEL, RMN & MODELISAT MOLEC LAB, F-67070 STRASBOURG, FRANCE
COUNTRY OF AUTHOR: USA; FRANCE
SOURCE: JOURNAL OF COMPUTATIONAL CHEMISTRY, (JAN/FEB 1992) Vol. 13, No. 1, pp. 1-11.
ISSN: 0192-8651.
DOCUMENT TYPE: Article; Journal
FILE SEGMENT: PHYS
LANGUAGE: ENGLISH
REFERENCE COUNT: 33
ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB An **algorithm** for the rapid analytical determination of the accessible surface areas of solute molecules is described. The accessible surface areas as well as the derivatives with respect to the Cartesian **coordinates** of the atoms are **computed** by a program called "MSEED," which is based in part on Connolly's analytical formulas for determining surface area. Comparisons of the CPU time required for MSEED, Connolly's numerical **algorithm** DOT, and a program for surface area determination (ANA) based on Connolly's analytical **algorithm,** are presented. MSEED is shown to be as much as 70 times faster than ANA and up to 11 times faster than DOT for several proteins. The greater speed of MSEED is achieved partially because nonproductive **computation** of the surface areas of internal atoms is avoided. A sample minimization of an **energy function,** which included a term for hydration, was carried out on MET-enkephalin using MSEED to **compute** the solvent-accessible surface area and its derivatives. The potential employed was ECEPP/2 plus an empirical potential for solvation based on the solvent-accessible surface area of the peptide. The CPU time required for 150 steps of minimization with the potential that included solvation was approximately twice as great as the CPU time required for 150 steps of minimization with the ECEPP/2 potential

only.

L166 ANSWER 36 OF 70 SCISEARCH COPYRIGHT 2003 THOMSON ISI on STN
ACCESSION NUMBER: 91:487295 SCISEARCH
THE GENUINE ARTICLE: GC251
TITLE: CONSTRAINED OPTIMIZATION AND **PROTEIN-STRUCTURE** DETERMINATION
AUTHOR: EISENFELD J; VAJDA S; SUGAR I; DELISI C (Reprint)
CORPORATE SOURCE: BOSTON UNIV, COLL ENGN, 44 CUMMINGTON ST, BOSTON, MA, 02215; BOSTON UNIV, DEPT BIOMED ENGN, BOSTON, MA, 02215; CUNY MT SINAI SCH MED, DEPT BIOMATH SCI, NEW YORK, NY, 10029
COUNTRY OF AUTHOR: USA
SOURCE: AMERICAN JOURNAL OF PHYSIOLOGY, (1991) Vol. 261, No. 2, pp. C376-C386.
DOCUMENT TYPE: Article; Journal
FILE SEGMENT: LIFE
LANGUAGE: ENGLISH
REFERENCE COUNT: 29
ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB Energy minimization is one of the main approaches to the **computational** determination of macromolecular structure. Due to the approximations in the empirical free-**energy functions** and due to the **computational** difficulties in locating their global minima, the problem is at present intractable when the only information available is the sequence of subunits forming the molecule. A less-demanding problem in terms of both physics and mathematics is constrained optimization, which uses additional but incomplete experimental information such as **distances** between certain **atoms**. This paper reviews methods for generating molecular structure using bond lengths and angles as variables and shows how the structure can be fully specified in terms of local geometry. The analysis permits precise statements to be made about the minimum set of distances that specify a unique structure without recourse to energy minimization. We then discuss the complementary situation, i.e., structure prediction with energy minimization based only on sequence information. Finally, we show how distance constraints can be incorporated into energy minimization methods.

L166 ANSWER 37 OF 70 CAPLUS COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER: 2001:834467 CAPLUS
DOCUMENT NUMBER: 136:17660
TITLE: Constrained global optimization for estimating molecular structure from **atomic distances**
AUTHOR(S): Williams, Glenn A.; Dugan, Jonathan M.; Altman, Russ B.
CORPORATE SOURCE: Stanford Medical Informatics, Stanford University, Stanford, CA, 94305-5479, USA
SOURCE: Journal of Computational Biology (2001), 8(5), 523-547 CODEN: JCOBEM; ISSN: 1066-5277
PUBLISHER: Mary Ann Liebert, Inc.
DOCUMENT TYPE: Journal
LANGUAGE: English

AB Finding optimal three-dimensional mol. configurations based on a limited amt. of exptl. and/or theor. data requires efficient nonlinear optimization algorithms. Optimization methods must be able to find at. configurations that are close to the abs., or global, min. error and also satisfy known phys. constraints such as min. sepn. distances between atoms (based on van der Waals interactions). The most difficult obstacles in these types of problems are that (1) using a limited amt. of input data leads to many possible local optima and (2) introducing phys. constraints, such as min. sepn. distances, helps to limit the search space but often

makes convergence to a global min. more difficult. We introduce a constrained global optimization algorithm that is robust and efficient in yielding near-optimal three-dimensional configurations that are guaranteed to satisfy known sepn. constraints. The algorithm uses an atom-based approach that reduces the dimensionality and allows for tractable enforcement of constraints while maintaining good global convergence properties. We evaluate the new optimization algorithm using synthetic data from the yeast phenylalanine tRNA and several proteins, all with known crystal structure taken from the Protein Data Bank. We compare the results to commonly applied optimization methods, such as distance geometry, simulated annealing, continuation, and smoothing. We show that compared to other optimization approaches, our algorithm is able combine sparse input data with phys. constraints in an efficient manner to yield structures with lower root mean squared deviation.

REFERENCE COUNT: 47 THERE ARE 47 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L166 ANSWER 38 OF 70 CAPLUS COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER: 2000:792196 CAPLUS
DOCUMENT NUMBER: 134:112591
TITLE: Extracting knowledge-based **energy functions** from protein structures by error rate minimization: Comparison of methods using lattice model
AUTHOR(S): Xia, Yu; Levitt, Michael
CORPORATE SOURCE: Department of Structural Biology, Stanford University School of Medicine, Stanford, CA, 94305, USA
SOURCE: Journal of Chemical Physics (2000), 113(20), 9318-9330
CODEN: JCPSA6; ISSN: 0021-9606
PUBLISHER: American Institute of Physics
DOCUMENT TYPE: Journal
LANGUAGE: English

AB We describe a general framework for extg. knowledge-based **energy function** from a set of native protein structures. In this scheme, the **energy function** is optimal when there is least chance that a random structure has a lower energy than the corresponding native structure. We first show that subject to certain approxns., most current database-derived **energy functions** fall within this framework, including **mean-field** potentials, Z-score optimization, and constraint satisfaction methods. We then propose a simple method for **energy function** parametrization derived from our anal. We go on to compare our method to other methods using a simple lattice model in the context of three different **energy function** scenarios. We show that our method, which is based on the most stringent criteria, performs best in all cases. The power and limitations of each method for deriving knowledge-based **energy function** is examd.

REFERENCE COUNT: 30 THERE ARE 30 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L166 ANSWER 39 OF 70 CAPLUS COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER: 2001:341191 CAPLUS
DOCUMENT NUMBER: 135:223677
TITLE: Identifying sequence-structure pairs undetected by sequence alignments
AUTHOR(S): Miyazawa, Sanzo
CORPORATE SOURCE: Faculty of Technology, Gunma University, Gunma, 376, Japan
SOURCE: Genome Informatics Series (2000), 11(Genome Informatics 2000), 141-150
CODEN: GINSE9; ISSN: 0919-9454
PUBLISHER: Universal Academy Press
DOCUMENT TYPE: Journal

LANGUAGE: English

AB A protein sequence-structure alignment method for database searches is examd. on how effectively this method together with a simple scoring function previously developed can identify compatibilities between sequences and structures of proteins. The scoring function consists of pairwise contact energies, repulsive packing potentials of residues for overly dense arrangement and short-range potentials for secondary structures. Pairwise contact interactions in a sequence-structure alignment are evaluated in a **mean field** approxn. on the basis of probabilities of site pairs to be aligned. Gap penalties are assumed to be proportional to the no. of contacts at each residue position, and as a result gaps will be more frequently placed on protein surfaces than in cores. In addn. to min. energy alignments, we use probability alignments made by successively aligning site pairs in order by pairwise alignment probabilities. Results show that the present **energy function** and alignment method can detect well both folds compatible with a given sequence and, inversely, sequences compatible with a given fold. Probability alignments consisting of most reliable site pairs only can yield small root mean square deviations, and including less reliable pairs increases the deviations. Remarkably, by this method some individual sequence-structure pairs are detected having only 5-20% sequence identity.

REFERENCE COUNT: 30 THERE ARE 30 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L166 ANSWER 40 OF 70 CAPLUS COPYRIGHT 2003 ACS on STN

ACCESSION NUMBER: 1996:515460 CAPLUS

DOCUMENT NUMBER: 125:242272

TITLE: From secondary structure to three-dimensional structure: improved dihedral angle probability distribution function for use with energy searches for native structures of polypeptides and proteins

AUTHOR(S): Cheng, Betty; Nayeem, Akbar; Scheraga, Harold A.

CORPORATE SOURCE: Baker Lab. Chem., Cornell Univ., Ithaca, NY, 14853-1301, USA

SOURCE: Journal of Computational Chemistry (1996), 17(12), 1453-1480

CODEN: JCCHDD; ISSN: 0192-8651

PUBLISHER: Wiley

DOCUMENT TYPE: Journal

LANGUAGE: English

AB An improved scheme to help in the prediction of protein structure is presented. This procedure generates improved starting conformations of a protein suitable for energy minimization. Trivariate Gaussian distribution functions for the .vphi., .psi., and .chi.1 dihedral angles have been derived, using conformational data from high resoln. protein structures selected from the Protein Data Bank (PDB). These trivariate probability functions generate initial values for the .vphi., .psi., and .chi.1 dihedral angles which reflect the exptl. space by focusing the search mainly in the regions of native proteins. The efficiency of the new trivariate probability distributions is demonstrated by comparing the results for the .alpha.-class polypeptide fragment, the mutant Antennapedia (C39 .fwdarw. S) homeodomain (2HOA), with those from two ref. probability functions. The first ref. probability function is a uniform or flat probability function and the second is a bivariate probability function for .vphi. and .psi.. The trivariate Gaussian probability functions are shown to search the conformational space more efficiently than the other two probability functions. The trivariate Gaussian probability functions are also tested on the binding domain of Streptococcal protein G (2GB1), an .alpha./.beta. class protein. Since presently available **energy functions** are not accurate enough to identify the most native-like energy-minimized structures, three selection criteria were used to identity a native-like structure with a

1.90-.ANG. rmsd from the NMR structure as the best structure for the Antennapedia fragment. Each individual selection criterion (ECEPP/3 energy, ECEPP/3 energy-plus-free energy of hydration, or a knowledge-based **mean field** method) was unable to identify a native-like structure, but simultaneous application of more than one selection criterion resulted in a successful identification of a native-like structure for the Antennapedia fragment. In addn. to these tests, structure predictions are made for the Antennapedia polypeptide, using a Pattern Recognition-based Importance-Sampling Minimization (PRISM) procedure to predict the backbone conformational state of the mutant Antennapedia homeodomain. The ten most probable backbone conformational state predictions were used with the trivariate and bivariate Gaussian dihedral angle probability distributions to generate starting structures (i.e., dihedral angles) suitable for energy minimization. The final energy-minimized structures show that neither the trivariate nor the bivariate Gaussian probability distributions are able to overcome the inaccuracies in the backbone conformational state predictions to produce a native-like structure. Until highly accurate predictions of the backbone conformational states become available, application of these dihedral angle probability distributions must be limited to problems, such as homol. modeling, in which only a limited portion of the backbone (e.g., surface loops) must be explored.

L166 ANSWER 41 OF 70 CAPLUS COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER: 1996:14623 CAPLUS
DOCUMENT NUMBER: 124:49452
TITLE: Inter-C.alpha. **atomic** potentials derived from the statistics of average interresidue **distances** in proteins: application to bovine pancreatic trypsin inhibitor
AUTHOR(S): Kikuchi, Takeshi
CORPORATE SOURCE: International Research Lab., Ciba-Geigy (Japan) Ltd., Takarazuka, 665, Japan
SOURCE: Journal of Computational Chemistry (1996), 17(2), 226-37
CODEN: JCCHDD; ISSN: 0192-8651
PUBLISHER: Wiley
DOCUMENT TYPE: Journal
LANGUAGE: English

AB New effective potentials acting between pairs of residues in proteins are proposed based on statistics of av. distances and std. deviations between C.alpha. atoms of residues in protein tertiary structures. Gaussian functions are adopted as anal. forms of the potentials. A protein structure is modeled as a chain mol. with a fixed bond length connecting particles approximating the effects of amino acid residues. The potentials derived in this study are used for conformational sampling of trypsin inhibitor from bovine pancreas. Sampling is done with the Monte Carlo simulated annealing method. Sampled conformations can be classified into a few groups or structural classes, and one of these classes contains structures relatively close (with 7.8-8.7 .ANG. root mean square [rms] deviation) to the x-ray structure. The native structure exhibits relatively low energy. These results denote a rather smooth landscape of the present potential energy surfaces. One class of classified structures contains native like structures, which suggests that the native structure can be predicted by further refinement of structures in this class. The authors discuss other properties and the effectiveness of the present potentials for description of protein structures.

L166 ANSWER 42 OF 70 CAPLUS COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER: 1995:317161 CAPLUS
DOCUMENT NUMBER: 122:126697
TITLE: Statistical thermodynamics of protein folding: comparison of a mean-field theory with Monte Carlo

simulations
AUTHOR(S): Hao, Ming-Hong; Scheraga, Harold A.
CORPORATE SOURCE: Baker Lab. Chem., Cornell Univ., Ithaca, NY, 14853-1301, USA
SOURCE: Journal of Chemical Physics (1995), 102(3), 1334-48
CODEN: JCPSA6; ISSN: 0021-9606
PUBLISHER: American Institute of Physics
DOCUMENT TYPE: Journal
LANGUAGE: English

AB A comparative study of protein folding with an anal. theory and computer simulations, resp., is reported. The theory is based on an improved mean-field formalism which, in addn. to the usual mean-field approxns., takes into account the distributions of energies in the subsets of conformational states. Sequence-specific properties of proteins are parametrized in the theory by 2 sets of **variables**, 1 for the energetics of **mean-field** interactions and 1 for the distribution of energies. Simulations are carried out on model polypeptides with different sequences, with different chain lengths, and with different interaction potentials, ranging from strong biases toward certain local chain states (bond angles and torsional angles) to complete absence of local conformational preferences. Theor. anal. of the simulation results for the model polypeptides revealed 3 different types of behavior in the folding transition from the statistical coiled state to the compact globular state; these included a cooperative 2-state transition, a continuous folding, and a glasslike transition. It was found that, with the fitted theor. parameters which were specific for each polypeptide under a different potential, the mean-field theory could describe the thermodn. properties and folding behavior of the different polypeptides accurately. By comparing the theor. descriptions with simulation results, the authors verified basic assumptions of the theory and, thereby, obtained new insights about the folding transitions of proteins. It was found that the cooperativity of the 1st-order folding transition of the model polypeptides was detd. mainly by long-range interactions, in particular the dipolar orientation; the local interactions (e.g., bond-angle and torsion-angle potentials) had only marginal effect on the cooperative characteristic of the folding, but had a large impact on the difference in energy between the folded lowest-energy structure and the unfolded conformations of a protein.

L166 ANSWER 43 OF 70 CAPLUS COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER: 1993:644879 CAPLUS
DOCUMENT NUMBER: 119:244879
TITLE: Protein conformations in aqueous solution calculated by **distance** constraints between **atoms** from NMR
AUTHOR(S): Lin, Donghai; Zhou, Zhe; Wu, Qinyi
CORPORATE SOURCE: Dep. Chem., Xiamen Univ., Xiamen, 361005, Peop. Rep. China
SOURCE: Bopuxue Zazhi (1992), 9(4), 337-45
CODEN: BOZAE2; ISSN: 1000-4556
DOCUMENT TYPE: Journal
LANGUAGE: Chinese

AB A method of detg. the three-dimensional structure for proteins in aq. soln. by a set of distance constraints between backbond atoms (mainly for H-H, obtained from NMR) was developed. In this method, only dihedral angles were selected as independent variables, a proper target function was minimized by local-to-globular optimization, thus these dihedral angles and the coordinates of six kinds of the backbone atoms (N, H, Ca, Ha, C', O) were calcd. The program DISNMA for this method was designed, and proved with the std. structure of BPTI. It requires less amt. of memory space and computing time and can be used for other larger mols.

L166 ANSWER 44 OF 70 CAPLUS COPYRIGHT 2003 ACS on STN

ACCESSION NUMBER: 1990:555078 CAPLUS
DOCUMENT NUMBER: 113:155078
TITLE: Optimal design of multipurpose batch plants. 2. A decomposition solution strategy
AUTHOR(S): Papageorgaki, Savoula; Reklaitis, Gintaras V.
CORPORATE SOURCE: Sch. Chem. Eng., Purdue Univ., West Lafayette, IN, 47907, USA
SOURCE: Industrial & Engineering Chemistry Research (1990), 29(10), 2062-73
CODEN: IECRED; ISSN: 0888-5885
DOCUMENT TYPE: Journal
LANGUAGE: English

AB A mixed integer nonlinear programming NMINLPE formulation for the optimum design of a multipurpose plant is given in part 1. The complexity of the model makes the problem computationally intractable for direct soln. by existing MINLP soln. techniques. Consequently, a decompn. strategy is presented that alternately solves a MILP master problem, which dets. the values of the **binary assignment variables** for fixed campaign lengths, and a NLP subproblem, which performs equipment sizing and dets. the values of the campaign lengths. The effectiveness of the decompn. procedure is demonstrated with a no. of test problems that are solved in reasonable computation times.

L166 ANSWER 45 OF 70 CAPLUS COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER: 1989:20676 CAPLUS
DOCUMENT NUMBER: 110:20676
TITLE: Determination of three-dimensional structures of proteins from interproton **distance** data by dynamical simulated annealing from a random array of **atoms**. Circumventing problems associated with folding
AUTHOR(S): Nilges, Michael; Clore, G. Marius; Gronenborn, Angela M.
CORPORATE SOURCE: Max-Planck-Inst. Biochem., Martinsried, D-8033, Fed. Rep. Ger.
SOURCE: FEBS Letters (1988), 239(1), 129-36
CODEN: FEBLAL; ISSN: 0014-5793
DOCUMENT TYPE: Journal
LANGUAGE: English

AB A new real space method, based on the principles of simulated annealing, is presented for detg. protein structures on the basis of interproton distance restraints derived from NMR data via nuclear Overhauser effect techniques. The method circumvents the folding problem assocd. with all real space methods described to date, by starting from a completely random array of atoms and introducing the force consts. for the covalent, interproton distance and repulsive van der Waals terms in the target function appropriately. The system is simulated at high temp. by solving Newton's equations of motion. As the values of all force consts. are very low during the early stages of the simulation, energy barriers between different folds of the protein can be overcome, and the global min. of the target function is reliably located. Further, because the atoms are initially only weakly coupled, they can move essentially independently to satisfy the restraints. The method is illustrated by using 2 examples of small proteins, namely crambin (46 residues) and potato carboxypeptidase inhibitor (39 residues).

L166 ANSWER 46 OF 70 CAPLUS COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER: 1970:494636 CAPLUS
DOCUMENT NUMBER: 73:94636
TITLE: Computer approaches to protein structure. I. Analysis of **atomic distances**
AUTHOR(S): Tometsko, Andrew M.
CORPORATE SOURCE: Sch. of Med. and Dent., Univ. of Rochester, Rochester,

NY, USA
SOURCE: Computers and Biomedical Research (1970), 3(3), 229-37
CODEN: CBMRB7; ISSN: 0010-4809
DOCUMENT TYPE: Journal
LANGUAGE: English
AB X-ray crystallography has yielded the detailed 3-dimensional geometry for a no. of proteins. In order to interpret the atomic interactions within a protein mol., it is usually necessary to construct mol. models. An other means of acquiring information concerning mol. structure would be to recast the x-ray data into a more meaningful form. Computer programs, which convert the coordinates of the atoms into a form that facilitates the study of preferred atomic interactions, are described. In this approach the local atomic environment for each atom of the protein is calcd. and printed in a manner that makes information of specific atomic interactions readily accessible.

ACCESSION NUMBER: 2002076001 EMBASE
TITLE: A method for optimizing potential-**energy functions** by a hierarchical design of the potential-energy landscape: Application to the UNRES force field.
AUTHOR: Liwo A.; Arlukowicz P.; Czaplewski C.; Oldziej S.; Pillardy J.; Scheraga H.A.
CORPORATE SOURCE: H.A. Scheraga, Baker Laboratory of Chemistry, Cornell University, Ithaca, NY 14853-1301, United States. hasS@cornell.edu
SOURCE: Proceedings of the National Academy of Sciences of the United States of America, (19 Feb 2002) 99/4 (1937-1942).
Refs: 27
ISSN: 0027-8424 CODEN: PNASA6
COUNTRY: United States
DOCUMENT TYPE: Journal; Article
FILE SEGMENT: 029 Clinical Biochemistry
LANGUAGE: English
SUMMARY LANGUAGE: English
AB A method for optimizing potential-**energy functions** of proteins is proposed. The method assumes a hierarchical structure of the energy landscape, which means that the energy decreases as the number of native-like elements in a structure increases, being lowest for structures from the native family and highest for structures with no native-like element. A level of the hierarchy is defined as a family of structures with the same number of native-like elements (or degree of native likeness). Optimization of a potential-**energy function** is aimed at achieving such a hierarchical structure of the energy landscape by forcing appropriate free-energy gaps between hierarchy levels to place their energies in ascending order. This procedure is different from methods developed thus far, in which the energy gap and/or the Z score between the native structure and all non-native structures are maximized, regardless of the degree of native likeness of the non-native structures. The advantage of this approach lies in reducing the number of structures with decreasing energy, which should ensure the searchability of the potential. The method was tested on two proteins, PDB ID codes 1FSD and 1IGD, with an off-lattice united-residue force field. For 1FSD, the search of the conformational space with the use of the conformational space annealing method and the newly optimized potential-**energy function** found the native structure very quickly, as opposed to the potential-**energy functions** obtained by former optimization methods. After even incomplete optimization, the force field obtained by using 1IGD located the native-like structures of two peptides, 1FSD and betanova (a designed three-stranded .beta.-sheet peptide), as the lowest-energy conformations, whereas for the 46-residue N-terminal fragment of staphylococcal protein A, the native-like conformation was the

second-lowest-energy conformation and had an energy 2 kcal/mol above that of the lowest-energy structure.

L166 ANSWER 48 OF 70 EMBASE COPYRIGHT 2003 ELSEVIER SCI. B.V. on STN
ACCESSION NUMBER: 2002015276 EMBASE
TITLE: Exploratory studies of ab initio protein structure prediction: Multiple copy simulated annealing, AMBER **energy functions**, and a generalized born/solvent accessibility solvation model.
AUTHOR: Liu Y.; Beveridge D.L.
CORPORATE SOURCE: D.L. Beveridge, Chemistry Department, Molecular Biophysics Program, Wesleyan University, Middletown, CT 06457, United States. dbeveridge@wesleyan.edu
SOURCE: Proteins: Structure, Function and Genetics, (1 Jan 2002) 46/1 (128-146).
Refs: 107
ISSN: 0887-3585 CODEN: PSFGEY
COUNTRY: United States
DOCUMENT TYPE: Journal; Article
FILE SEGMENT: 029 Clinical Biochemistry
LANGUAGE: English
SUMMARY LANGUAGE: English

AB A theoretical and computational approach to ab initio structure prediction for polypeptides in water is described and applied to selected amino acid sequences for testing and preliminary validation. The method builds systematically on the extensive efforts applied to parameterization of molecular dynamics (MD) force fields, employs an empirically well-validated continuum dielectric model for solvation, and an eminently parallelizable approach to conformational search. The effective free energy of polypeptide chains is estimated from AMBER united atom potential functions, with internal degrees of freedom for both backbone and amino acid side chains explicitly treated. The hydration free energy of each structure is determined using the Generalized Born/Solvent Accessibility (GBSA) method, modified and reparameterized to include atom types consistent with the AMBER force field. The conformational search procedure employs a multiple copy, Monte Carlo simulated annealing (MCSA) protocol in full torsion angle space, applied iteratively on sets of structures of progressively lower free energy until a prediction of a structure with lowest effective free energy is obtained. Calibration tests for the effective **energy function** and search algorithm are performed on the alanine dipeptide, selected protein crystal structures, and united atom decoys on barnase, crambin, and six examples from the Rosetta set. Specific demonstration cases of the method are provided for the 8-mer sequence of Ala residues, a 12-residue peptide with longer side chains QLLKKLLQQLKQ, a de novo designed 16 residue peptide of sequence (AAQAA)3(3)Y, a 15-residue sequence with a .beta. sheet motif, GEWTWDATKTFTVTE, and a 36 residue small protein, Villin headpiece. The Ala 8-mer readily formed an .alpha.-helix. An .alpha.-helix structure was predicted for the 16-mer, consistent with observed results from IR and CD spectroscopy and with the pattern in .psi..phi. angles of known protein structures. The predicted structure for the 12-mer, composed of a mix of helix and less regular elements of secondary structure, lies 2.65 ARMS from the observed crystal structure. Structure prediction for the 8-mer .beta.-motif resulted in form 4.50 ARMS from the crystal geometry. For Villin, the predicted native form is very close to the crystal structure, RMS values of 3.5 A (including sidechains), and 1.01 A (main chain only). The methodology permits a detailed analysis of the molecular forces which dominate various segments of the predicted folding trajectory. Analysis of the results in terms of internal torsional, electrostatic and van der Waals and the electrostatic and non-electrostatic contributions to hydration, including the hydrophobic effect, is presented. .COPYRGT. 2001 Wiley-Liss, Inc.

L166 ANSWER 49 OF 70 EMBASE COPYRIGHT 2003 ELSEVIER SCI. B.V. on STN

ACCESSION NUMBER: 2001170537 EMBASE
TITLE: STRAP: Editor for STRuctural alignments of proteins.
AUTHOR: Gille C.; Frommel C.
CORPORATE SOURCE: C. Gille, Institute of Biochemistry, Medical Facul. of the Humboldt Univ., Monbijoustrasse 2a, 10117 Berlin, Germany
SOURCE: Bioinformatics, (2001) 17/4 (377-378).
Refs: 11
ISSN: 1367-4803 CODEN: BOINFP
COUNTRY: United Kingdom
DOCUMENT TYPE: Journal; Article
FILE SEGMENT: 027 Biophysics, Bioengineering and Medical Instrumentation
029 Clinical Biochemistry
LANGUAGE: English
SUMMARY LANGUAGE: English

AB STRAP is a comfortable and extensible tool for the generation and refinement of multiple alignments of protein sequences. Various sequence ordered input file formats are supported. These are the SwissProt-, GenBank-, EMBL-, DSSP- PDB-, MSF-, and plain ASCII text format. The special feature of STRAP is the simple visualization of spatial **distances** of C(.alpha.)-**atoms** within the alignment. Thus structural information can easily be incorporated into the sequence alignment and can guide the alignment process in cases of low sequence similarities. Further STRAP is able to manage huge alignments comprising a lot of sequences. The protein viewers and modeling programs INSIGHT, RASMOL and WEBMOL are embedded into STRAP. STRAP is written in Java. The well-documented source code can be adapted easily to special requirements. STRAP may become the basis for complex alignment tools in the future.

L166 ANSWER 50 OF 70 EMBASE COPYRIGHT 2003 ELSEVIER SCI. B.V. on STN

ACCESSION NUMBER: 2000208885 EMBASE
TITLE: On the design and analysis of protein folding potentials.
AUTHOR: Tobi D.; Shafran G.; Linial N.; Elber R.
CORPORATE SOURCE: R. Elber, Department of Computer Science, Cornell University, 4130 Upson Hall, Ithaca, NY 14853, United States. ron@cs.cornell.edu
SOURCE: Proteins: Structure, Function and Genetics, (1 Jul 2000) 40/1 (71-85).
Refs: 23
ISSN: 0887-3585 CODEN: PSFGEY
COUNTRY: United States
DOCUMENT TYPE: Journal; Article
FILE SEGMENT: 029 Clinical Biochemistry
LANGUAGE: English
SUMMARY LANGUAGE: English

AB Pairwise interaction models to recognize native folds are designed and analyzed. Different sets of parameters are considered but the focus was on 20 x 20 contact matrices. Simultaneous solution of inequalities and minimization of the variance of the energy find matrices that recognize exactly the native folds of 572 sequences and structures from the protein data bank (PDB). The set includes many homologous pairs, which present a difficult recognition problem. Significant recognition ability is recovered with a small number of parameters (e.g., the H/P model). However, full recognition requires a complete set of amino acids. In addition to structures from the PDB, a folding program (MONSSTER) was used to generate decoy structures for 75 proteins. It is impossible to recognize all the native structures of the extended set by contact potentials. We therefore searched for a new functional form. An **energy function** U, which is based on a sum of general pairwise interactions limited to a resolution of 1 angstrom, is considered. This set was infeasible too. We therefore conjecture that it is not possible to find a folding potential, resolved to 1 angstrom, which

is a sum of pair interactions. (C) 2000 Wiley-Liss, Inc.

L166 ANSWER 51 OF 70 EMBASE COPYRIGHT 2003 ELSEVIER SCI. B.V. on STN
ACCESSION NUMBER: 1998000296 EMBASE
TITLE: **Binary assignments** of amino acids from pattern conservation.
AUTHOR: Irback A.; Potthast F.
CORPORATE SOURCE: A. Irback, Department of Theoretical Physics, University of Lund, Solvegatan 14A, S-223 62 Lund, Sweden
SOURCE: Protein Engineering, (1997) 10/9 (1013-1017).
Refs: 16
ISSN: 0269-2139 CODEN: PRENE
COUNTRY: United Kingdom
DOCUMENT TYPE: Journal; Article
FILE SEGMENT: 029 Clinical Biochemistry
LANGUAGE: English
SUMMARY LANGUAGE: English
AB We have developed a simple optimization procedure for assigning binary values to amino acids. The binary values are determined by a maximization of the degree of pattern conservation in groups of closely related protein sequences. The maximization is carried out at fixed composition. For compositions approximately corresponding to an equipartition of the residues, the optimal encoding is found to be strongly correlated with hydrophobicity. The stability of the procedure is demonstrated. Our calculations are based upon sequences in the SWISS-PROT database.

L166 ANSWER 52 OF 70 EMBASE COPYRIGHT 2003 ELSEVIER SCI. B.V. on STN
ACCESSION NUMBER: 96178842 EMBASE
DOCUMENT NUMBER: 1996178842
TITLE: Self-consistently optimized statistical mechanical **energy functions** for sequence structure alignment.
AUTHOR: Koretke K.K.; Luthey-Schulten Z.; Wolynes P.G.
CORPORATE SOURCE: School of Chemical Sciences, University of Illinois,Urbana, IL 61801; United States
SOURCE: Protein Science, (1996) 5/6 (1043-1059).
ISSN: 0961-8368 CODEN: PRCIEI
COUNTRY: United States
DOCUMENT TYPE: Journal; Article
FILE SEGMENT: 029 Clinical Biochemistry
LANGUAGE: English
SUMMARY LANGUAGE: English
AB A quantitative form of the principle of minimal frustration is used to obtain from a database analysis statistical mechanical **energy functions** and gap parameters for aligning sequences to three-dimensional structures. The analysis that partially takes into account correlations in the energy landscape improves upon the previous approximations of Goldstein et al. (1994, 1995) (Goldstein R, Luthey-Schulten Z, Wolynes P, 1994, Proceedings of the 27th Hawaii International Conference on System Sciences. Los Alamitos, California: IEEE Computer Society Press. pp 306-315; Goldstein R, Luthey- Schulten Z, Wolynes P, 1995, In: Elber R, ed. New developments in theoretical studies of proteins. Singapore: World Scientific). The **energy function** allows for ordering of alignments based on the compatibility of a sequence to be in a given structure (i.e., lowest energy) and therefore removes the necessity of using percent identity or similarity as scoring parameters. The alignments produced by the **energy function** on distant homologues with low percent identity (less than 21%) are generally better than those generated with evolutionary information. The lowest energy alignment generated with the **energy function** for sequences containing prosite signatures but unknown structures is a structure containing the same prosite signature, providing a check on the robustness of the algorithm.

Finally, the **energy function** can make use of known experimental evidence as constraints within the alignment algorithm to aid in finding the correct structural alignment.

L166 ANSWER 53 OF 70 EMBASE COPYRIGHT 2003 ELSEVIER SCI. B.V. on STN
ACCESSION NUMBER: 96054787 EMBASE
DOCUMENT NUMBER: 1996054787
TITLE: Folding proteins with a simple **energy function** and extensive conformational searching.
AUTHOR: Yue K.; Dill K.A.
CORPORATE SOURCE: Dept. of Pharmaceutical Chemistry, University of California, Box 1204,San Francisco, CA 94143, United States
SOURCE: Protein Science, (1996) 5/2 (254-261).
ISSN: 0961-8368 CODEN: PRCIEI
COUNTRY: United States
DOCUMENT TYPE: Journal; Article
FILE SEGMENT: 027 Biophysics, Bioengineering and Medical Instrumentation
029 Clinical Biochemistry
LANGUAGE: English
SUMMARY LANGUAGE: English

AB We describe a computer algorithm for predicting the three-dimensional structures of proteins using only their amino acid sequences. The method differs from others in two ways: (1) it uses very few energy parameters, representing hydrophobic and polar interactions, and (2) it uses a new 'constraint-based exhaustive' searching method, which appears to be among the fastest and most complete search methods yet available for realistic protein models. It finds a relatively small number of low-energy conformations, among which are native-like conformations, for crambin (1CRN), avian pancreatic polypeptide (1PPT), melittin (2MLT), and apamin. Thus, the lowest-energy states of very simple **energy functions** may predict the native structures of globular proteins.

L166 ANSWER 54 OF 70 EMBASE COPYRIGHT 2003 ELSEVIER SCI. B.V. on STN
ACCESSION NUMBER: 95184432 EMBASE
DOCUMENT NUMBER: 1995184432
TITLE: An evaluation of discrete and continuum search techniques for conformational analysis of side chains in proteins.
AUTHOR: Vasquez M.
CORPORATE SOURCE: Protein Design Labs, Inc., 2375 Garcia Avenue,Mountain View, CA 94043, United States
SOURCE: Biopolymers, (1995) 36/1 (53-70).
ISSN: 0006-3525 CODEN: BIPMAA
COUNTRY: United States
DOCUMENT TYPE: Journal; Article
FILE SEGMENT: 029 Clinical Biochemistry
LANGUAGE: English
SUMMARY LANGUAGE: English

AB Methodology for calculation of side-chain conformations in proteins is evaluated. The role and impact of corrections to idealized rotameric structures are considered, by incorporating methods for torsional optimization into rotamer-packing algorithms. Off-rotamer corrections given by continuum torsional optimization improve, over simpler rotamer-packing procedures, the accuracy with which the conformations of side chains of buried amino acids can be predicted. The analogy between protein side-chain calculations and spin systems is explored by adapting spin simulation methods to side-chain packing algorithms. Implementations of **mean-field** and heat- bath algorithms for side-chain packing are described and their performance tested. The procedures introduced here address the combinatorial problem in an efficient and reasonably effective manner, as evidenced by analysis of their convergence properties. Application of refined protocols yields overall prediction accuracies of 80% for AHp1 and 68% for AHp1,2 pairs for a test set of 60

proteins, using a 40.degree. cutoff to define correct placement. For buried amino acids (defined as having less than 30% relative solvent accessibility) the prediction accuracies increase to 88% for AHp1 and 79% for AHp1,2 pairs. The influence of the form of the potential **energy function** is studied by comparing results obtained with 12-6 and 9-6 potentials. The 9-6 form leads to more accurate results. Detailed comparison with previous work is presented, and the effect of combinatorial packing steps is shown to be important for all but the smallest proteins.

L166 ANSWER 55 OF 70 WPIDS COPYRIGHT 2003 THOMSON DERWENT on STN
ACCESSION NUMBER: 2003-430565 [40] WPIDS
DOC. NO. NON-CPI: N2003-343679
DOC. NO. CPI: C2003-113938
TITLE: Producing an optimized pharmacopore for a target protein e.g. integral membrane **protein**, by using datasets comprising chemical **structure** information for several compounds and quantified properties for each compound.
DERWENT CLASS: B04 D16 S03 T01
INVENTOR(S): PENG, J W; VAN DRIE, J H
PATENT ASSIGNEE(S): (VERT-N) VERTEX PHARM INC
COUNTRY COUNT: 101
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 2003038442	A2	20030508	(200340)*	EN	47

RW: AT BE BG CH CY CZ DE DK EA EE ES FI FR GB GH GM GR IE IT KE LS LU MC MW MZ NL OA PT SD SE SK SL SZ TR TZ UG ZM ZW
W: AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ OM PH PL PT RO RU SD SE SG SI SK SL TJ TM TN TR TT TZ UA UG US UZ VC VN YU ZA ZM ZW

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 2003038442	A2	WO 2002-US34512	20021029

PRIORITY APPLN. INFO: US 2001-350080P 20011029
AB WO2003038442 A UPAB: 20030624
NOVELTY - Producing (M1) an optimized pharmacopore, comprises selecting a first dataset comprising chemical structure information of several compounds and quantified property of each of the compounds, applying a **computational** unit to the first dataset to generate a first pharmacophore, applying a second **computational** unit to a second dataset to produce the optimized pharmacophore and outputting the optimized pharmacophore to a suitable output device.

DETAILED DESCRIPTION - Producing (M1) an optimized pharmacopore, comprises:

(a) selecting a first dataset comprising chemical structure information of several compounds and a first quantified property of each of the compounds, where the property is related to the affinity of each of the compounds to a target protein;

(b) applying a first **computational** unit to a first dataset to generate a first pharmacophore;

(c) applying a second **computational** unit to a second dataset to produce the optimized pharmacophore, where the second dataset comprises one, two or all of the first pharmacopore, the first data set

and the first quantified property, and a second quantified property for each of the compounds, where the property is related to the conformation of each of the compounds when it is bound to the target protein; and

(d) outputting the optimized pharmacophore to a suitable output device.

INDEPENDENT CLAIMS are also included for:

(1) a process (M2) for identifying a compound having an affinity to a target protein, by selecting an optimized pharmacophore for the target protein, virtually screening each of the several molecular structures in a database against the optimized pharmacophore to identify a molecular structure having structural features that substantially satisfy structural constraints of the optimized pharmacophore, and outputting the molecular structure to a suitable output device;

(2) identifying (M3) a compound structure having an affinity to a target protein, by selecting an optimized pharmacophore for the target **protein**, identifying a discrete **structure** element corresponding to each structural constraint of the optimized pharmacophore and creating with it a molecular scaffold, mining the scaffold to identify a molecular structure having structural features that substantially satisfy structural constraints of the optimized pharmacophore, and outputting the molecular structure to a suitable output device;

(3) designing a ligand for a target protein, by identifying a compound whose molecular structure substantially satisfies structural constraints of an optimized pharmacophore for the target protein;

(4) a **computer** for designing a ligand for a target protein, comprising a machine-readable data storage medium comprising a data storage material encoded with machine-readable data, where the data comprises, an optimized pharmacophore, and several molecular structures, a working memory for storing a **computational** unit for processing the machine-readable data, a central-processing unit coupled to the working memory and to the machine-readable data storage medium for processing the machine- readable data to identify a molecular structure using the instructions, and an output device coupled to the central-processing unit for outputting the results; and

(5) a process for identifying optimized compounds.

USE - (M1) is useful for producing an optimized pharmacophore. (M2) is useful for identifying a compound having affinity to a target protein, and (M3) is useful for identifying a compound structure having an affinity to a target protein. The target protein is an integral membrane protein, a membrane- tethered protein, preferably G-protein coupled receptor (GPCR), ion-channel proteins, transporter proteins or cytokine receptors (claimed).

Dwg.0/0

L166 ANSWER 56 OF 70 WPIDS COPYRIGHT 2003 THOMSON DERWENT on STN

ACCESSION NUMBER:	2003-092925 [08] WPIDS
DOC. NO. NON-CPI:	N2003-073764
DOC. NO. CPI:	C2003-023182
TITLE:	Inhibiting human methionine aminopeptidase 2, by administering compounds with certain structural, and physical characteristics that allow for interacting compounds with specific residues of active site of enzyme.
DERWENT CLASS:	B03 D16 T01
INVENTOR(S):	MARINO, J P; RYAN, M D; SMITH, W W; THOMPSON, S K
PATENT ASSIGNEE(S):	(SMIK) SMITHKLINE BEECHAM CORP
COUNTRY COUNT:	100
PATENT INFORMATION:	

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 2002081415	A2	20021017	(200308)*	EN	789

RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW MZ

NL OA PT SD SE SL SZ TR TZ UG ZM ZW
W: AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK
DM DZ EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR
KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ OM PH PL PT
RO RU SD SE SG SI SK SL TJ TM TN TR TT TZ UA UG US UZ VN YU ZA ZM
ZW

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 2002081415	A2	WO 2002-US9458	20020328

PRIORITY APPLN. INFO: US 2001-281221P 20010403

AB WO 200281415 A UPAB: 20030204

NOVELTY - Inhibiting (M1) human methionine aminopeptidase 2 (hMetAP2), involves administering compounds with certain structural, physical, and spatial characteristics that allow for the interaction of compounds with specific residues of the active site of the enzyme.

DETAILED DESCRIPTION - Inhibiting (M1) human methionine aminopeptidase 2 (hMetAP2) involves, either:

(A) administering to a mammal in need of a compound comprising one or two heteroatoms that fits spatially into the active site of hMetAP2, the compound comprises any one of the following characteristics:

(i) an interaction, singly, or jointly as a pair, to one or both metals in the active site of hMetAP2, where the metals are selected from cobalt, zinc, manganese, iron and nickel, and where the heteroatoms are 1.5-3.5 Angstrom from the closest metal;

(ii) hydrogen bonding interactions with histidine 231;

(iii) hydrophobic interactions with atoms of two or more amino acid residues selected from tyrosine 444, histidine 231, histidine 382, alanine 413, tyrosine 383, phenylalanine 219, proline 220, methionine 384, glycine 222, and isoleucine 338;

(iv) hydrophobic interactions with one or more residue selected from histidine 339, isoleucine 338 or tyrosine 444;

(v) hydrogen bonding interactions with asparagine 315; and

(vi) hydrophobic interactions with one or more residues selected from leucine 328, leucine 447, histidine 231, or alanine 230; or

(B) administering to a mammal in need of a compound comprising one or two heteroatoms that fits spatially into the active site of hMetAP2, the compound comprising any one of the following:

(i) an interaction, singly, or jointly as a pair, to one or both metals in the active site of hMetAP2, where the heteroatom is 1.7-2.4 Angstrom from the closest metal;

(ii) an atom capable of hydrogen bonding such as oxygen, nitrogen, or sulfur, that interacts with histidine 231, such that the **distance** between the **atom** and histidine 231 is 2.2-4.5 Angstrom ;

(iii) a hydrophobic group that interacts with residue selected from tyrosine 444, histidine 231, histidine 382, alanine 413, tyrosine 383, phenylalanine 219, proline 220, methionine 384, glycine 222, and isoleucine 338, such that the distance between the hydrophobic group atom and the residue is 2.2-4.5 Angstrom ;

(iv) a hydrophobic group that interacts with isoleucine 338, tyrosine 444, or histidine 339, where histidine 339 may exist in one of at least three different conformations selected from the pairs of side chain rotameric angles chi-1 and chi-2: (-164,177), (-150,-133), or (-70,-149);

(v) a group capable of hydrogen bonding such as carbonyl oxygen or an ether oxygen that interacts with asparagine 315; and

(vi) a hydrophobic group such as methyl group in contact with leucine 328, leucine 447, histidine 231, or alanine 230, such that the distance between the hydrophobic group and the residue is 3.4-5.0 Angstrom .

INDEPENDENT CLAIMS are also included for the following:

(1) identifying (M2) an inhibitor compound capable of binding to, and inhibiting the proteolytic activity of, hMetAP2, involves:

(i) introducing into a suitable **computer** program information defining an active site conformation of hMetAP2 molecule comprising a catalytically active site, the active site defined by the protein coordinates fully given in the specification, where the program displays its three-dimensional structure,

(ii) creating a three-dimensional representation of the active site cavity in the **computer** program,

(iii) displaying and superimposing the model of the test compound on the model of the active site,

(iv) assessing whether the test compound model fits spatially into the active site,

(v) preparing the test compound that fits spatially into the active site,

(vi) using the test compound in a biological assay for a protease characterized by the active site, and

(vii) determining whether the test compound inhibits hMetAP2 activity in the assay;

(2) a peptide, peptidomimetic or synthetic molecule identified by M2;

(3) designing (M3) drug, involves using structure coordinates of a hMetAP2 crystal to **computationally** evaluate a chemical entity of associating with the active site of hMetAP2;

(4) identifying (M4) inhibitors which competitively bind to the active site of a hMetAP2 molecule or its fragment characterized by a catalytically active site, the active site defined by the protein coordinates, involves:

(i) providing the coordinates of the active site of the protease to a **computerized** modeling system,

(ii) identifying compounds which will bind to the structure, and

(iii) screening the compounds identified for protease inhibitory bioactivity; and

(5) identifying (M5) a potential inhibitor for a hMetAP2 enzyme, involves:

(i) using a three-dimensional structure of the enzyme as defined by the protein coordinates,

(ii) employing the three-dimensional structure to design or select the potential inhibitor,

(iii) synthesizing the potential inhibitor, and

(iv) contacting the potential inhibitor with the enzyme in the presence of a substrate to determine the ability of the potential inhibitor to inhibit the enzyme.

ACTIVITY - Cytostatic; Antirheumatic; Antiarthritic; Antiatherosclerotic; Antipsoriatic; Anorectic; Ophthalmological.

MECHANISM OF ACTION - Inhibitor of hMetAP2. No biological data is given.

USE - M1 is useful for inhibiting hMetAP2 (claimed), and the compounds which are administered to inhibit hMetAP2 is useful for treating conditions mediated by angiogenesis, such as cancer, hemangioma, proliferative retinopathy, rheumatoid arthritis, atherosclerotic neovascularization, psoriasis, ocular neovascularization and obesity.
Dwg.0/21

L166 ANSWER 57 OF 70 WPIDS

ACCESSION NUMBER:	2003-029954 [02] WPIDS
DOC. NO. NON-CPI:	N2003-023700
DOC. NO. CPI:	C2003-006840
TITLE:	New method using a computer for quantitative protein design and automation by applying a protein design cycle to each of the inputted protein backbone scaffolds and generating a probability matrix derived from variable sequences.
DERWENT CLASS:	B04 D16 T01

```
INVENTOR(S):            DESJARLAIS, J R
PATENT ASSIGNEE(S):     (DESJ-I) DESJARLAIS J R; (PENN-N) PENN STATE RES FOUND
COUNTRY COUNT:          100
PATENT INFORMATION:

     PATENT NO     KIND DATE     WEEK      LA   PG
     ---------------------------------------------
     WO 2002077751 A2 20021003 (200302)* EN   90
        RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW MZ
            NL OA PT SD SE SL SZ TR TZ UG ZM ZW
         W: AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK
            DM DZ EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR
            KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ OM PH PL PT
            RO RU SD SE SG SI SK SL TJ TM TN TR TT TZ UA UG UZ VN YU ZA ZM ZW
     US 2002147547 A1 20021010 (200302)
     US 2003036854 A1 20030220 (200316)

APPLICATION DETAILS:

     PATENT NO     KIND                  APPLICATION      DATE
     ----------------------------------------------------------
     WO 2002077751 A2                    WO 2002-US3789   20020206
     US 2002147547 A1 Provisional        US 2001-266711P  20010206
                                         US 2001-877695   20010608
     US 2003036854 A1 Provisional        US 2001-266711P  20010206
                      Cont of            US 2001-877695   20010608
                                         US 2002-71859    20020206


PRIORITY APPLN. INFO: US 2001-877695   20010608; US 2001-266711P
                      20010206; US 2002-71859    20020206
```

AB WO 200277751 A UPAB: 20030111

NOVELTY - A method executed by a computer under the control of a program comprising inputting an ensemble of protein backbone scaffolds, applying at least one protein design cycle to each of the scaffolds, and generating a probability matrix derived from variable sequences, is new. The computer includes a memory for storing the program.

DETAILED DESCRIPTION - An INDEPENDENT CLAIM is also included for a method for optimizing simulation or scoring function parameters that uses comparisons between designed sequences and natural sequences.

USE - The method is useful for quantitative protein design and automation.

ADVANTAGE - The method reduces the number of wasted sequences produced in the experimental library and reduces the cost and difficulty of protein engineering. Further, the method allows analysis of multiple backbone states, rather than just one, to sample an even larger amount of possible amino acid sequence space.

Dwg.0/8

```
L166 ANSWER 58 OF 70  WPIDS  COPYRIGHT 2003 THOMSON DERWENT on STN
ACCESSION NUMBER:      2001-648409 [74]   WPIDS
DOC. NO. NON-CPI:      N2001-484498
DOC. NO. CPI:          C2001-191341
TITLE:                 Computer implemented method for predicting
                       binding sites of ligands in proteins, uses annealing
                       molecular dynamics including solvation effects in
                       optimizing preferred binding conformations.
DERWENT CLASS:         B04 D16 S03
INVENTOR(S):           FLORIANO, W B; GODDARD, W A; VAIDEHI, N
PATENT ASSIGNEE(S):    (CALY) CALIFORNIA INST OF TECHNOLOGY; (FLOR-I) FLORIANO W
                       B; (GODD-I) GODDARD W A; (VAID-I) VAIDEHI N
COUNTRY COUNT:         96
PATENT INFORMATION:
```

```
PATENT NO      KIND DATE     WEEK        LA    PG
--------------------------------------------------
WO 2001071347 A1 20010927 (200174)* EN    72
   RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW MZ
       NL OA PT SD SE SL SZ TR TZ UG ZW
    W: AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK
       DM DZ EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ
       LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PL PT RO RU SD
       SE SG SI SK SL TJ TM TR TT TZ UA UG US UZ VN YU ZA ZW
AU 2001050955 A  20011003 (200210)
US 2002099506 A1 20020725 (200254)
EP 1272839    A1 20030108 (200311)   EN
    R: AL AT BE CH CY DE DK ES FI FR GB GR IE IT LI LT LU LV MC MK NL PT
       RO SE SI TR
```

APPLICATION DETAILS:

PATENT NO	KIND		APPLICATION	DATE
WO 2001071347	A1		WO 2001-US9402	20010323
AU 2001050955	A		AU 2001-50955	20010323
US 2002099506	A1	Provisional	US 2000-191895P	20000323
		Provisional	US 2000-213658P	20000623
		Cont of	US 2001-816772	20010323
			US 2001-10725	20011130
EP 1272839	A1		EP 2001-924288	20010323
			WO 2001-US9402	20010323

FILING DETAILS:

PATENT NO	KIND		PATENT NO
AU 2001050955	A	Based on	WO 200171347
EP 1272839	A1	Based on	WO 200171347

PRIORITY APPLN. INFO: US 2000-213658P 20000623; US 2000-191895P 20000323; US 2001-816772 20010323; US 2001-10725 20011130

AB WO 200171347 A UPAB: 20011217

NOVELTY - A **computer** implemented method uses structural information for protein to identify a binding region. Preferred binding conformations are identified for each set of ligand. The conformations are optimized using annealing molecular dynamics including solvation effects. The lowest binding energy calculated is selected, and output as the predicted binding energies for each set of ligand.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the following:

(1) A similar method for predicting the **structure** of a **protein** binding site for a protein with an unknown binding site;

(2) A **computational** model of a ligand-protein complex for a protein having an unknown binding site comprising a **computer** -readable memory storing data describing an optimized preferred binding **conformation** for the **protein** and a ligand known to bind to the protein;

(3) A **computer** program product on a **computer** -readable medium for modeling ligand-protein interaction comprising instructions to operate the method;

(4) A **computer**-implemented method of generating a pharmacophore comprising performing the method and generating a pharmacophore model based on the optimized binding conformations and outputting the design.

USE - The **computer** implemented method is useful in

applications, e.g. performing fast screening of virtual chemical compound libraries against targets of pharmacological interest, fast scanning of globular and membrane bound proteins for potential binding sites, prediction of potential ligands and ligand binding modes, and prediction of receptor function based on selective binding affinities. It can also be used in identifying the interaction of cellular receptors with surface structures expressed by microbial pathogens to understand the molecular basis of pathogenesis.

ADVANTAGE - The invention provides **computationally** efficient and accurate models for predicting binding site of ligands in proteins and drug design.

DESCRIPTION OF DRAWING(S) - Figure shows a flow diagram illustrating a general **computational** protocol for modeling ligand-protein interactions according to the method.

Dwg.1/14

L166 ANSWER 59 OF 70 WPIDS COPYRIGHT 2003 THOMSON DERWENT on STN
ACCESSION NUMBER: 2001-343922 [36] WPIDS
DOC. NO. NON-CPI: N2001-249058
DOC. NO. CPI: C2001-106545
TITLE: Analyzing a **protein structure** by systematical analysis in terms of individual contributions of single, pairs and occasionally also multiplets of amino acid residues to the global energy of a protein, useful for designing proteins.
DERWENT CLASS: B04 D16 T01
INVENTOR(S): BOUTTON, C; DESMET, J; LASTERS, I; STAS, P; VLIEGHE, D
PATENT ASSIGNEE(S): (ALGO-N) ALGONOMICS NV
COUNTRY COUNT: 95
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 2001037147	A2	20010525	(200136)*	EN	96

RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW MZ NL OA PT SD SE SL SZ TR TZ UG ZW
W: AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CR CU CZ DE DK DM DZ EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT TZ UA UG US UZ VN YU ZA ZW

PATENT NO	KIND	DATE	WEEK	LA	PG
AU 2001021554	A	20010530	(200152)		
EP 1242925	A2	20020925	(200271)	EN	

R: AL AT BE CH CY DE DK ES FI FR GB GR IE IT LI LT LU LV MC MK NL PT RO SE SI

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 2001037147	A2	WO 2000-EP10923	20001103
AU 2001021554	A	AU 2001-21554	20001103
EP 1242925	A2	EP 2000-984970	20001103
		WO 2000-EP10923	20001103

FILING DETAILS:

PATENT NO	KIND		PATENT NO
AU 2001021554	A	Based on	WO 200137147
EP 1242925	A2	Based on	WO 200137147

PRIORITY APPLN. INFO: US 1999-163409P 19991103
AB WO 200137147 A UPAB: 20010628

NOVELTY - A new method (M1) for analyzing a **protein structure** comprises the systematical analysis of known **protein structures** in terms of individual contributions of single, pairs and occasionally also multiplets of amino acid residues to the global energy of a protein comprising any of these residues.

DETAILED DESCRIPTION - A new method (M1) for analyzing a **protein structure** comprises the systematical analysis of known **protein structures** in terms of individual contributions of single, pairs and occasionally also multiplets of amino acid residues to the global energy of a protein comprising any of these residues.

In detail, M1 is executable in a **computer** under the control of a program stored in the **computer**, and comprises:

(a) receiving a reference **structure** for a **protein** , where the reference **structure** forms a representation of a 3D **structure** of the **protein** which consists of many residue positions, each carrying a particular reference amino acid type in a specific reference **conformation**, and the **protein** residues are classified into a set of modeled residue positions and a set of fixed residues, the latter being included into a fixed template;

(b) substituting into the reference structure of step (a) a pattern which consists of one or more of the modeled residue positions defined in step (a), each carrying a particular amino acid residue type in a fixed conformation, and the amino acid residue types of the pattern are replacing the corresponding amino acid residue types present in the reference structure;

(c) optimizing the global conformation of the reference structure of step (a) being substituted by the pattern of step (b), where:

(i) a suitable **protein structure** optimization method based on a function allowing to assess the quality of a global **protein structure**, or any part of it, is used in combination with a suitable conformational search method;

(ii) the structure optimization method is applied to all modeled residue positions defined in step (a) not being located at any of the pattern residue positions defined in step (b); and

(iii) the pattern and template residues are kept fixed;

(d) assessing the energetic compatibility (EC) of the pattern defined in step (b) within the context of the reference structure defined in step (a) being structurally optimized in step (a) with respect to the pattern, by way of comparing the global energy of the substituted and optimized **protein structure** with the global energy of the non-substituted reference structure; and

(e) storing a value reflecting the EC of the pattern together with information related to the structure of the pattern in the form of an energetic compatibility object (ECO).

INDEPENDENT CLAIMS are also included for the following:

(1) a fold recognition method to identify a potential structural relationship between a particular target amino acid sequence and one or more **protein** 3D **structures**, the **protein** 3D **structures** being analyzed by M1;

(2) an inverse folding method to identify a potential structural relationship between a particular **protein** 3D **structure** and one or more known amino acid sequences, where the **protein** 3D **structure** is analyzed by M1;

(3) a protein design method to identify or generate amino acid sequences which are energetically compatible with a particular **protein** 3D **structure**, where the **protein** 3D **structure** is analyzed by M1;

(4) a type-dependent, topology-specific solvation method (M2) for the assignment of a set of energetic solvation terms to a set of residue types, depending on the degree of solvent exposure of their respective rotamers at the considered residue positions in a **protein structure**;

(5) a nucleic acid sequence (N1) encoding a protein sequence analyzed by M1;
(6) an expression vector comprising N1;
(7) a host cell comprising (N1);
(8) a method of treating a disease in a mammal, comprising administering a pharmaceutical composition comprising a therapeutically effective amount of a protein sequence analyzed by M1;
(9) an ECO obtainable by M1;
(10) a database in the form of a data structure comprising a set of ECO's obtainable by M1;
(11) a **computing** device for analyzing a **protein structure**, comprising a means for carrying out steps (a) to (d) of M1, and a memory for storing a value reflecting the EC of the pattern together with information related to the structure of the pattern in the form of an ECO as a data structure;
(12) a **computing** device for carrying out M2;
(13) a **computer** program product to be utilized for **computing** on a **computing** system with a processor and memory, comprising instruction means for carrying out steps (a) to (e) of M1 or for carrying out M2;
(14) a **computer** readable data carrier comprising an executable **computer** program product of (13) or for executing any of the above methods; and
(15) a method comprising:
(a) a description of a least a **protein** reference **structure** at a near location;
(b) transmitting the description to a remote processing engine running a **computer** program for carrying out any of the above methods; and
(c) receiving at a near location from the remote processing engine an output of the above methods.

USE - The method is useful for analyzing known protein structures by computing a quantitative measure reflecting the energetic compatibility of all naturally occurring and synthetic residues of interest at each residue position of interest in the structure. Therefore, the method is useful for designing proteins. The protein analyzed by M1 can be used for treating a disease in a mammal.

ADVANTAGE - The method preserves the accuracy of the most accurate atom-based modelling techniques currently known while avoiding the bottle-neck problem known as the combinatorial substitution problem, therefore gaining several orders of magnitude in computational speed.
Dwg.0/5

L166 ANSWER 60 OF 70 WPIDS COPYRIGHT 2003 THOMSON DERWENT on STN

ACCESSION NUMBER:	2001-432565 [46] WPIDS
DOC. NO. NON-CPI:	N2001-320549
DOC. NO. CPI:	C2001-130835
TITLE:	A novel method for identifying the interaction site, binding site or active site in a macromolecule, using of informative combinatorial chemistry, informative peptide libraries and Multivariate Quantitative Structure Activity Relationships.
DERWENT CLASS:	A96 B04 D16 J04 S03
INVENTOR(S):	ANDERSSON, P; LUNDSTEDT, T; MUCENIECE, R; PRUSIS, P; WIKBERG, J
PATENT ASSIGNEE(S):	(MELA-N) MELACURE THERAPEUTICS AB; (PETT-I) PETT C P
COUNTRY COUNT:	95
PATENT INFORMATION:	

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 2001036980	A2	20010525	(200146)*	EN	100

RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW MZ

```
          NL OA PT SD SE SL SZ TR TZ UG ZW
       W: AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CR CU CZ DE DK DM
          DZ EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC
          LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PL PT RO RU SD SE
          SG SI SK SL TJ TM TR TT TZ UA UG US UZ VN YU ZA ZW
   AU 2001015305 A  20010530 (200152)
   EP 1232466    A2 20020821 (200262)  EN
       R: AL AT BE CH CY DE DK ES FI FR GB GR IE IT LI LT LU LV MC MK NL PT
          RO SE SI TR
```

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 2001036980	A2	WO 2000-GB4420	20001120
AU 2001015305	A	AU 2001-15305	20001120
EP 1232466	A2	EP 2000-977666	20001120
		WO 2000-GB4420	20001120

FILING DETAILS:

PATENT NO	KIND		PATENT NO
AU 2001015305	A	Based on	WO 200136980
EP 1232466	A2	Based on	WO 200136980

PRIORITY APPLN. INFO: GB 1999-27346 19991118

AB WO 200136980 A UPAB: 20010815

NOVELTY - Characterizing (M1) the interaction between a Ligand Y and a Target X by obtaining information representing one or more physical and/or chemical properties of targets of type X and type Y to produce a model of interaction

DETAILED DESCRIPTION - Characterizing the interaction between a Ligand Y and a Target X comprising:

(a) the following steps:

(i) obtaining information representing one or more chemical and/or physical properties of at least two ligands of the type Y;

(ii) obtaining information representing one or more chemical and/or physical properties of at least two targets of type X; and

(iii) obtaining information representing one or more chemical and/or physical properties of the interactions between at least two of the ligands of type Y and at least two of the ligands of type X; and

(b) processing the information from (i), (ii) and (iii) to produce a model of the interaction between the Ligand Y and the Target X from which one or more properties of the interaction between the Ligand Y and the Target X may be identified and/or characterized.

INDEPENDENT CLAIMS are also included for the following:

(1) estimating (M2) the position of the active site in a Target X in an interaction between a Ligand Y and a Target X, or estimating one or more physical and/or chemical properties of the active site, comprising:

(a) the steps (i)-(iii) in M1; and

(b) correlating the information from (i)-(iii) to produce a model of the interaction between the Ligand Y and the Target X from which the position of the active site or one or more physical and/or chemical properties of the active site in the Target X may be estimated;

(2) identifying (M3) the position of the active site in an interaction between a Ligand Y and a Target X, or predicting one or more physical and/or chemical properties of the active site, comprising:

(a) the steps (i)-(iii) in M1;

(b) correlating the information from (i)-(iii) to produce a model of the interaction between the Ligand Y and the Target X; and

(c) using the model to identify the position of the active site or one or more physical and/or chemical properties of the active site.

(3) a process (M4) performed with the aid of a programmed **computer** for the estimation of the position of the active site in a target X, in an interaction between a Ligand Y and a Target X, or one or more physical and/or chemical properties of the active site, comprising:

(a) the steps of:

(i) inputting information representing one or more chemical and/or physical properties of at least two ligands of the type Y;

(ii) inputting information representing one or more chemical and/or physical properties of at least two ligands of the type X;

(iii) inputting information representing one or more chemical and/or physical properties of the interaction between at least two of the ligands of type Y and at least two of the targets of the type X;

(iv) **computing** or calculating a model from the inputted information which describes the interaction between the ligand Y and the Target X; and

(b) using the model to estimate the position of the active site and/or to estimate one or more physical and/or chemical properties of the active site.

(4) a process (M5) for assisting in the design of a Ligand Y' which binds to a Target X, the Ligand Y' having an increased or decreased binding affinity, selectivity or avidity for the Target X compared to that of a Ligand Y, comprising:

(a) the steps of:

(i) the steps (i)-(iii) in M1; and

(ii) correlating the informing from (i)-(iii) to produce a model of the interaction between the Ligand Y and the Target X from which the structure and/or one or more chemical and/or physical properties of the Ligand Y' may be estimated or predicted;

(5) estimating or predicting (M6) the binding affinity, selectivity or avidity of a Ligand Y' with a Target X, comprising:

(a) steps (i)-(iii) of M1; and

(b) correlating the information from steps (i)-(iii) to produce a model of the interaction between the Ligand Y and the Target X from which the binding affinity, selectivity or avidity of the Ligand Y' may be estimated or predicted;

(6) producing (M7) a Ligand Y' which binds to a Target X, the Ligand Y' having an increased or decreased binding affinity, selectivity or activity for the target X compared to that of a Ligand Y, comprising:

(a) steps (i)-(iii) of M1;

(b) correlating the informing from (i)-(iii) to produce a model of the interaction between the Ligand Y and the Target X from which the structure and/or one or more properties of the Ligand Y' may be estimated or predicted; and

(c) producing the Ligand Y' by a method known per se;

(7) a lead, organic compound, catalyst, pharmaceutical, drug, macromolecule being capable of binding a molecule, peptide, peptidomimetic, protein, enzyme, antibody, molecule, macromolecule, DNA, RNA, carbohydrate when designed by a process comprising any of M1-M5;

(8) computer software specifically adapted to carry out the processes given in the specification, when installed on data processing means; and

(9) a ligand whose structure and or properties has been estimated or predicted through the use of a process claimed in the specification.

USE - The methods of the invention are useful for:

(a) identifying outliers of type X or outliers of type Y;

(b) drug design;

(c) design or identification of lead compounds;

(d) design of ligands of type Y with improved affinity and/or selectivity for targets of type X;

(e) protein engineering;

(f) design of DNA or RNA molecules;

(g) design of artificial targets of type X/or artificial ligands of type Y;

(h) analysis and/or in the engineering of regions and/or parts of targets of type X and/or ligands of type Y;

(i) design of organic compound, catalyst, pharmaceutical, drug, macromolecule being capable of binding a molecule, peptidomemetic, protein, enzyme, antibody, molecule, macromolecule, DNA, RNA or a carbohydrate;

(j) the design of a ligand of type Y being capable of binding a target of type X;

(k) design of any one of organic compound, catalyst, pharmaceutical, drug, macromolecule capable of binding a molecule, peptide, peptidomemetic, protein, enzyme, antibody, molecule and a macromolecule; and

(l) designing new ligands for known targets and/or for new targets.

Dwg.0/30

```
L166 ANSWER 61 OF 70  WPIDS  COPYRIGHT 2003 THOMSON DERWENT on STN
ACCESSION NUMBER:     2001-273356 [28]   WPIDS
DOC. NO. NON-CPI:     N2001-195272
DOC. NO. CPI:         C2001-082833
TITLE:                Optimizing specific building blocks which make up a
                      target macromolecule involves determining several
                      conformers of each building block, which are quantified
                      and ranked using scoring function and reference
                      structure.
DERWENT CLASS:        B04 D16 T01
INVENTOR(S):          LACROIX, E; SERRANO, L
PATENT ASSIGNEE(S):   (EUMO-N) EURO MOLECULAR BIOLOGY LAB; (LACR-I) LACROIX E;
                      (SERR-I) SERRANO L
COUNTRY COUNT:        94
PATENT INFORMATION:

     PATENT NO    KIND DATE     WEEK      LA   PG
     ---------------------------------------------
     WO 2001016810 A2 20010308 (200128)* EN  138
        RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW MZ
            NL OA PT SD SE SL SZ TZ UG ZW
         W: AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CR CU CZ DE DK DM
            DZ EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC
            LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PL PT RO RU SD SE
            SG SI SK SL TJ TM TR TT TZ UA UG US UZ VN YU ZA ZW
     AU 2001011320 A  20010326 (200137)
     US 2002072864 A1 20020613 (200243)

APPLICATION DETAILS:

     PATENT NO    KIND                   APPLICATION      DATE
     ----------------------------------------------------------
     WO 2001016810 A2                    WO 2000-EP8504   20000831
     AU 2001011320 A                     AU 2001-11320    20000831
     US 2002072864 A1                    US 1999-387741   19990831

FILING DETAILS:

     PATENT NO    KIND                   PATENT NO
     ---------------------------------------------
     AU 2001011320 A  Based on           WO 200116810

PRIORITY APPLN. INFO: US 1999-387741    19990831
AB   WO 200116810 A UPAB: 20010522
```

NOVELTY - Choosing set of substitute building blocks (SBB) for set of positions in target macromolecule by determining conformations or conformers (I) of each produced SBB, minimizing calculated energy value by adjusting geometry of each (I) to obtain solution structure (ST),

calculating solution score (SS) having entropic term, for ST and choosing specified set of SBB if calculated SS is lower than a threshold value.

DETAILED DESCRIPTION - Choosing a set of SBB for a set of positions in target macromolecule according to whether a calculated SS is lower than a threshold value (M1) involves:

(a) specifying at least one SBB for each position in the set of positions to produce a specified set of SBB,

(b) for each SBB determining at least one (I), substituting coordinates of each (I) or its portion for coordinates of the building blocks or its portion at the position in an atomic structure of the target macromolecule;

(c) minimizing the value of a calculated energy term by adjusting the geometry of each (I) or its portion in order to obtain ST,

(d) calculating a SS for ST, in which SS comprises an entropic term, and

(e) choosing the specified set of SBB if the calculated SS is lower than a threshold value.

INDEPENDENT CLAIMS are also included for the following:

(1) a computer program product (II) for use in conjunction with a computer, the computer program product comprising a computer readable storage medium and a computer program mechanism embedded in it, the computer program mechanism comprising an optimizer module configured to choose a set of SBB for a set of positions in a target macromolecule according to whether a calculated SS is lower than a threshold value, the computer program mechanism, upon receiving as input the set of positions, performs (M1); and

(2) a system (III) for choosing a set of SBB for a set of positions in a target macromolecule according to whether a calculated SS is lower than a threshold value comprising a central processing unit, an input device for inputting requests, an output device, a memory, at least one bus connected to the central processing unit, the memory, the input device, and the output device, the memory storing an computer program comprising an optimizer module configured to choose the set of SBBs, the computer program mechanism, upon receiving a request to choose the set of SBB, performs (M1).

USE - Choosing a set of SBB for a set of positions in target macromolecule according to whether a calculated SS is lower than a threshold value (claimed). (M1) is useful for engineering and designing molecules which comprise building blocks that are individually amenable to systemic variation. The technique has applications in designing development of macromolecules for e.g. proteins, peptides, nucleic acids and polymers with desired properties.

ADVANTAGE - The novel method for designing and engineering macromolecules utilizes an accurate and complete mathematical representation of macromolecule structure, in order to reliably predict how precise variants of its sequence can be accommodated into a desired three dimensional structure.

DESCRIPTION OF DRAWING(S) - The figure shows the block diagram of a computer system.

Dwg.1/17

L166 ANSWER 62 OF 70	WPIDS COPYRIGHT 2003 THOMSON DERWENT on STN
ACCESSION NUMBER:	2000-038624 [03] WPIDS
DOC. NO. NON-CPI:	N2000-029163
DOC. NO. CPI:	C2000-009847
TITLE:	Three dimensional **computer** modeling to identify HLA binding compounds useful for modulating immune responses.
DERWENT CLASS:	B02 B05 D16 T01
INVENTOR(S):	KOHLER, N; LIU, M; RICHERT, J R; WANG, S; WU, X; YIN, D
PATENT ASSIGNEE(S):	(GEOU) UNIV GEORGETOWN; (KOHL-I) KOHLER N; (LIUM-I) LIU M; (RICH-I) RICHERT J R; (WANG-I) WANG S; (WUXX-I) WU X; (YIND-I) YIN D

COUNTRY COUNT: 87
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 9955682	A1	19991104	(200003)*	EN	64

RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW NL OA PT SD SE SL SZ UG ZW
W: AE AL AM AT AU AZ BA BB BG BR BY CA CH CN CU CZ DE DK EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT UA UG UZ VN YU ZA ZW

PATENT NO	KIND	DATE	WEEK	LA	PG
AU 9936691	A	19991116	(200015)		
EP 1082310	A1	20010314	(200116)	EN	

R: AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE

PATENT NO	KIND	DATE	WEEK	LA	PG
US 2002042423	A1	20020411	(200227)		
JP 2002513008	W	20020508	(200234)		59

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 9955682	A1	WO 1999-US9218	19990429
AU 9936691	A	AU 1999-36691	19990429
EP 1082310	A1	EP 1999-918878	19990429
		WO 1999-US9218	19990429
US 2002042423	A1 Provisional	US 1998-83426P	19980429
		US 1999-301339	19990429
JP 2002513008	W	WO 1999-US9218	19990429
		JP 2000-545842	19990429

FILING DETAILS:

PATENT NO	KIND		PATENT NO
AU 9936691	A	Based on	WO 9955682
EP 1082310	A1	Based on	WO 9955682
JP 2002513008	W	Based on	WO 9955682

PRIORITY APPLN. INFO: US 1998-83426P 19980429; US 1999-301339 19990429

AB WO 9955682 A UPAB: 20000118
NOVELTY - Methods (A) for identifying and using HLA (Histocompatability Lymphocyte-A System) binding compounds as HLA-agonists and antagonists, comprising the **computational** processing of a database containing 3-dimensional structures of receptor sites and chemical compounds, are new.

DETAILED DESCRIPTION - Method (A) of processing a compound data base containing 3-dimensional structures of chemical compounds to provide a lead compound capable of blocking a receptor site in a host molecule comprising:

(1) modeling the 3-dimensional structure of the receptor site;

(2) positioning a compound from the data base in the receptor site and assigning a geometrical-fit score, indicating the fit between the compound and the receptor site;

(3) ranking the compounds in the data base according to their score, and forming a group of compounds with a rank of a predetermined value, or higher;

(4) minimizing an **energy function** describing interactions between a compound and a receptor site by adjusting coordinates of the compound to obtain a minimum energy compound-host molecule complex structure;

(5) ranking the compounds according to their minimum energy values

and forming a subgroup of compounds with a minimum-energy rank of a predetermined value or higher; and

(6) visualizing (on a **computer**) a minimum energy compound-host molecule complex and forming a second subgroup of compounds with a visual-fit satisfying a predetermined criterion.

An INDEPENDENT CLAIM is also included for a method of inhibiting the interaction of an HLA molecule to an antigen comprising the administration of at least one compound of formula (I).

R1, R2 = optionally substituted phenyl, benzyl or other 5- or 6-membered aromatic ring system, optionally containing one or more heteroatoms selected from O, S and N

R3, R4 = H, optionally substituted phenyl, benzyl or other aromatic ring system, 1-10 C alkyl, 1-10 C alkoxy, halogen, SO3M, amide, or COOR
M = H or alkyl
R1 = H or alkyl

R5, R6, R7, R8 = H, halogen (F, Cl, Br, I), alkyl, 1-10C alkoxy, amide, NO2, amine, 1-10C cycloalkyl, nitroso, OH, ether, ester, sulfonic acid, alkenyl or allyl
X, Y = N or C

ACTIVITY - Immunosuppresive.

MECHANISM OF ACTION - HLA-agonists and antagonists.

USE - Compounds identified by (A), or of formula (I) can be used to treat autoimmune diseases, graft versus host disease, transplant rejection and multiple sclerosis.

ADVANTAGE - (A) allows adjusting of a compound's structure to optimize the fit between the host molecule and homologous compound. The method also allows for the modeling of host **proteins** whose 3-D **structure** is unknown.

DESCRIPTION OF DRAWING(S) - The diagram shows a 3-D model of the HLA-DR301 molecule produced by homology modeling.
Dwg.1/10

L166 ANSWER 63 OF 70 WPIDS COPYRIGHT 2003 THOMSON DERWENT on STN
ACCESSION NUMBER: 1997-021413 [02] WPIDS
DOC. NO. NON-CPI: N1997-017696
TITLE: **Computer** implemented method predicting 3D **protein** fragment **structure** - gives **protein** fragment 3D **structure** from amino acid sequence, hierarchic procedure staged ascends folding hierarchy with concomitant structure accretion represents fragment by geometry, folds under primitive **energy function** influence.
DERWENT CLASS: T01
INVENTOR(S): ROSE, G D; SRINIVASAN, R
PATENT ASSIGNEE(S): (UYJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE; (UYJO) UNIV JOHNS HOPKINS SCHOOL MED
COUNTRY COUNT: 21
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 9637856	A1	19961128	(199702)*	EN	42
RW: AT BE CH DE DK ES FI FR GB GR IE IT LU MC NL PT SE					
W: AU CA JP					
AU 9659327	A	19961211	(199713)		
US 5680319	A	19971021	(199748)		25

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 9637856	A1	WO 1996-US7713	19960523
AU 9659327	A	AU 1996-59327	19960523

US 5680319 A US 1995-450390 19950525

FILING DETAILS:

PATENT NO	KIND		PATENT NO
AU 9659327	A	Based on	WO 9637856

PRIORITY APPLN. INFO: US 1995-450390 19950525
AB WO 9637856 A UPAB: 19970108

The **computer** implemented method (32, 34, 36 and 38) creates an ensemble of conformations satisfying a localised energy condition using a random selection strategy. Persistently stable segments of the fragment in the ensemble are identified. A next similar ensemble is created with a larger locality window.

The ensemble is also constrained not to change the conformation types of persistently stable segments. Further similar segments are identified. The creation and identification are repeated for increasingly large locality windows. A conformation in the last formed ensemble is outputted from the **computer** as the predicted 3D structure.

USE - Is **computer** assisted method for determining 3D **structure** of **proteins**.

ADVANTAGE - Predicts fold of protein from its amino acid sequence alone.
Dwg.3/8

L166 ANSWER 64 OF 70 WPIDS COPYRIGHT 2003 THOMSON DERWENT on STN
ACCESSION NUMBER: 1994-358467 [44] WPIDS
CROSS REFERENCE: 1996-433261 [43]; 1998-052548 [05]
DOC. NO. NON-CPI: N1994-280825
DOC. NO. CPI: C1994-163629
TITLE: **Computer** modelling of three dimensional **protein structure** - based on **structure** of template **protein**, suitable for variable regions of proteins with weak sequence identity.
DERWENT CLASS: B04 J04 J05 S03 T01
INVENTOR(S): SRINIVASAN, S; SUDARSANAM, P
PATENT ASSIGNEE(S): (IMMV) IMMUNEX CORP
COUNTRY COUNT: 24
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 9425860	A1	19941110	(199444)*	EN	52
RW: AT BE CH DE DK ES FR GB GR IE IT LU MC NL PT SE					
W: AU CA FI JP KR NO NZ					
AU 9467799	A	19941121	(199508)		
US 5453937	A	19950926	(199544)		16

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 9425860	A1	WO 1994-US4822	19940428
AU 9467799	A	AU 1994-67799	19940428
US 5453937	A	US 1993-55050	19930428

FILING DETAILS:

PATENT NO	KIND		PATENT NO
AU 9467799	A	Based on	WO 9425860

PRIORITY APPLN. INFO: US 1993-55050 19930428
AB WO 9425860 A UPAB: 19990511

Computer modelling of the 30 **structure** of a model **protein** (MP) is based on the 30 structure of a template (TP). First for each amino acid (AA) in MP, when MP and TP have aligned AA, the position of each backbone atom in AA is established based on that of the topologically equivalent atom (TEA) in the aligned AA of TP. Then interatomic distance constraints for each pair of atoms with established positions are generated, and the position of each atom in MP is set to be within these constraints. Opt. (a) the conformance of the 3D **structure** with the rules of **protein** folding is assessed to identify a TP for a family of related proteins or (b) the method is applied to several different sequence alignments to give accurate sequence alignment between MP and TP.

ADVANTAGE - This method produces structures with minimum short contacts. It can be applied where there is only weak sequence identity between TP and MP, and to model variable regions positioned between 2 conserved regions.

Dwg.1/10

L166 ANSWER 65 OF 70 USPATFULL on STN
ACCESSION NUMBER: 2003:189033 USPATFULL
TITLE: Protein design automation for protein libraries
INVENTOR(S): Bentzien, Joerg, White Plains, NY, UNITED STATES
Dahiyat, Bassil I., Altadena, CA, UNITED STATES
Desjarlais, John R., Pasadena, CA, UNITED STATES
Hayes, Robert J., Pasadena, CA, UNITED STATES
Vielmetter, Jost, Altadena, CA, UNITED STATES

	NUMBER	KIND	DATE	
PATENT INFORMATION:	US 2003130827	A1	20030710	
APPLICATION INFO.:	US 2002-218102	A1	20020812	(10)

RELATED APPLN. INFO.: Continuation of Ser. No. US 2001-927790, filed on 10 Aug 2001, PENDING

	NUMBER	DATE
PRIORITY INFORMATION:	US 2001-311545P	20010810 (60)
	US 2001-324899P	20010925 (60)
	US 2002-351937P	20020125 (60)
	US 2002-352103P	20020125 (60)

DOCUMENT TYPE: Utility
FILE SEGMENT: APPLICATION
LEGAL REPRESENTATIVE: ROBIN M. SILVA, DORSEY & WHITNEY LLP, SUITE 3400, FOUR EMBARCADERO CENTER, SAN FRANCISCO, CA, 94111
NUMBER OF CLAIMS: 116
EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 29 Drawing Page(s)
LINE COUNT: 5782
CAS INDEXING IS AVAILABLE FOR THIS PATENT.

AB The invention relates to the use of protein design automation (PDA.TM.) to generate computationally prescreened secondary libraries of proteins, and to methods and compositions utilizing the libraries.

L166 ANSWER 66 OF 70 USPATFULL on STN
ACCESSION NUMBER: 2003:189003 USPATFULL
TITLE: Protein modeling tools
INVENTOR(S): Skolnick, Jeffrey, Creve Corner, MD, UNITED STATES
Kolinski, Andrzej, Warsaw, POLAND

NUMBER KIND DATE

```
                              ---------------- ------ --------
PATENT INFORMATION:           US 2003130797      A1    20030710
APPLICATION INFO.:            US 2001-982488     A1    20011017  (9)
RELATED APPLN. INFO.:         Division of Ser. No. US 2000-493022, filed on 27 Jan
                              2000, PENDING

                                   NUMBER         DATE
                              ------------------ --------
PRIORITY INFORMATION:         US 1999-117570P    19990127 (60)
                              US 1999-118844P    19990205 (60)
DOCUMENT TYPE:                Utility
FILE SEGMENT:                 APPLICATION
LEGAL REPRESENTATIVE:         GREGORY P. EINHORN, Fish & Richardson P.C., Suite 500,
                              4350 La Jolla Village Drive, San Diego, CA, 92122
NUMBER OF CLAIMS:             13
EXEMPLARY CLAIM:              1
NUMBER OF DRAWINGS:           21 Drawing Page(s)
LINE COUNT:                   4029
```

AB The invention provides a new, efficient method for the assembly of **protein** tertiary **structure** from known, loosely encoded secondary structure constraints and sparse information about exact side chain contacts. The method is based on a new method for the reduced modeling of **protein structure** and dynamics, where the **protein** is described by representing side chain centers of mass rather than alpha-carbons. The model has implicit, built-in multi-body correlations that simulate short- and long-range packing preferences, hydrogen bonding cooperativity, and a mean force potential describing hydrophobic interactions. Due to the simplicity of the protein representation and definition of the model force field, the Monte Carlo algorithm is at least an order of magnitude faster than previously published Monte Carlo algorithms for three-dimensional structure assembly. In contrast to existing algorithms, the new method requires a smaller number of tertiary constraints for successful fold assembly; on average, one for every seven residues as compared to one for every four residues. The reliability and robustness of the invention make it useful for routine application in model building protocols based on various (and even very sparse) experimentally-derived structural constraints.

```
L166 ANSWER 67 OF 70  USPATFULL on STN
ACCESSION NUMBER:             2003:51894  USPATFULL
TITLE:                        Apparatus and method for designing proteins and protein
                              libraries
INVENTOR(S):                  Desjarlais, John R., Pasadena, CA, UNITED STATES
PATENT ASSIGNEE(S):           The Penn State Research Foundation (U.S. corporation)

                                   NUMBER        KIND    DATE
                              ---------------- ------ --------
PATENT INFORMATION:           US 2003036854      A1    20030220
APPLICATION INFO.:            US 2002-71859      A1    20020206  (10)
RELATED APPLN. INFO.:         Continuation of Ser. No. US 2001-877695, filed on 8 Jun
                              2001, PENDING

                                   NUMBER         DATE
                              ------------------ --------
PRIORITY INFORMATION:         US 2001-266711P    20010206 (60)
DOCUMENT TYPE:                Utility
FILE SEGMENT:                 APPLICATION
LEGAL REPRESENTATIVE:         PAUL D. GREELEY, ESQ., OHLANDT, GREELEY, RUGGIERO &
                              PERLE, L.L.P., 10th FLOOR, ONE LANDMARK SQUARE,
                              STAMFORD, CT, 06901-2682
NUMBER OF CLAIMS:             17
EXEMPLARY CLAIM:              1
```

NUMBER OF DRAWINGS: 6 Drawing Page(s)
LINE COUNT: 2411
AB Methodology executed by a computer under the control of a program, said computer including a memory for storing said program, said method comprising the steps of inputting an ensemble of protein backbone scaffolds; applying at least one protein design cycle to each of said scaffolds; and generating a probability matrix derived from a plurality of variable sequences.

L166 ANSWER 68 OF 70 USPATFULL on STN
ACCESSION NUMBER: 2002:266659 USPATFULL
TITLE: Apparatus and method for designing proteins and protein libraries
INVENTOR(S): Desjarlais, John R., State College, PA, UNITED STATES

	NUMBER	KIND	DATE	
PATENT INFORMATION:	US 2002147547	A1	20021010	
APPLICATION INFO.:	US 2001-877695	A1	20010608	(9)

	NUMBER	DATE
PRIORITY INFORMATION:	US 2001-266711P	20010206 (60)

DOCUMENT TYPE: Utility
FILE SEGMENT: APPLICATION
LEGAL REPRESENTATIVE: Paul D. Greeley, Esq., Ohlandt, Greeley, Ruggiero & Perle, L.L.P., One Landmark Square, 10th Floor, Stamford, CT, 06901-2682
NUMBER OF CLAIMS: 37
EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 6 Drawing Page(s)
LINE COUNT: 1204
CAS INDEXING IS AVAILABLE FOR THIS PATENT.
AB Methodology for the automated design of proteins is disclosed. Various methods executed by a computer for generating probability matrices, protein sequences, combinatorial libraries of proteins, and optimization of various parameters related to protein design are disclosed. Methodology is applicable to the design and analysis of **protein structures** and **protein** sequences.

L166 ANSWER 69 OF 70 USPATFULL on STN
ACCESSION NUMBER: 2002:141873 USPATFULL
TITLE: COMPUTER-BASED METHOD FOR MACROMOLECULAR ENGINEERING AND DESIGN
INVENTOR(S): LACROIX, EMMANUEL, LEIMEN, GERMANY, FEDERAL REPUBLIC OF
SERRANO, LUIS, HEIDELBERG, GERMANY, FEDERAL REPUBLIC OF

	NUMBER	KIND	DATE	
PATENT INFORMATION:	US 2002072864	A1	20020613	
APPLICATION INFO.:	US 1999-387741	A1	19990831	(9)

DOCUMENT TYPE: Utility
FILE SEGMENT: APPLICATION
LEGAL REPRESENTATIVE: PENNIE AND EDMONDS, 1155 AVENUE OF THE AMERICAS, NEW YORK, NY, 100362711
NUMBER OF CLAIMS: 160
EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 10 Drawing Page(s)
LINE COUNT: 2673
AB The present invention relates to a system and method for engineering and designing a macromolecule. An experimentally determined or de novo atomic structure that corresponds to the macromolecule is identified. The atomic structure is composed of building blocks. When the

macromolecule is a peptide or a protein, the building blocks are amino acid residues. A target subset of the building blocks in the atomic structure to be optimized is identified. The coordinates of those building blocks that are not in the target subset are fixed. For each building block in the target subset, a large number of potential conformers is sample d. Each conformer to be sampled is substituted into the atomic structure and tested against an **energy function** that includes the equivalent energy of the conformer in a reference state. Combinations of conformers that best satisfy an interaction **energy function** are identified.

```
L166 ANSWER 70 OF 70  USPATFULL on STN
ACCESSION NUMBER:       2001:97645  USPATFULL
TITLE:                  Three dimensional structure of a ZAP tyrosine protein
                        kinase fragment and modeling methods
INVENTOR(S):            Hatada, Marcos H., Charlestown, MA, United States
                        Lu, Xiaode, Revere, MA, United States
                        Laird, Ellen R., Newton, MA, United States
                        Karas, Jennifer L., Lexington, MA, United States
                        Zoller, Mark J., Weston, MA, United States
                        Holt, Dennis A., Stow, MA, United States
PATENT ASSIGNEE(S):     Ariad Pharmaceuticals, Inc., Cambridge, MA, United
                        States (U.S. corporation)

                             NUMBER        KIND    DATE
                        ---------------- ------ --------
PATENT INFORMATION:     US 6251620         B1   20010626
APPLICATION INFO.:      US 1997-975040          19971118  (8)
RELATED APPLN. INFO.:   Division of Ser. No. US 1996-605578, filed on 22 Feb
                        1996

                               NUMBER        DATE
                        ------------------ --------
PRIORITY INFORMATION:   US 1995-3312P      19950906 (60)
DOCUMENT TYPE:          Utility
FILE SEGMENT:           GRANTED
PRIMARY EXAMINER:       Nashed, Nashaat T.
LEGAL REPRESENTATIVE:   Vincent, Matthew P., Clauss, Isabelle M.Foley, Hoag &
                        Eliot
NUMBER OF CLAIMS:       7
EXEMPLARY CLAIM:        1
NUMBER OF DRAWINGS:     15 Drawing Figure(s); 14 Drawing Page(s)
LINE COUNT:             19810
CAS INDEXING IS AVAILABLE FOR THIS PATENT.
```

AB The invention relates to human ZAP-70, and in particular, to the region of ZAP-70 containing the tandem Src homology-2 ("SH2") domains, to crystalline forms thereof, liganded or unliganded, which are particularly useful for the determination of the three-dimensional structure of the protein. The three dimensional structure of the tandem SH2 region of ZAP provides information useful for the design of pharmaceutical compositions which inhibit the biological function of ZAP and other members of the ZAP family of SH2 domain-containing proteins, particularly those biological functions mediated by molecular interactions involving one or both SH2 domains.

FILE 'HOME' ENTERED AT 12:13:09 ON 15 AUG 2003

```
Connecting via Winsock to STN



Welcome to STN International!  Enter x:x

LOGINID:SSSPTA1635SXZ

PASSWORD:
TERMINAL (ENTER 1, 2, 3, OR ?):2


* * * * * * * * * *   Welcome to STN International   * * * * * * * * * *

 NEWS  1          Web Page URLs for STN Seminar Schedule - N. America
 NEWS  2          "Ask CAS" for self-help around the clock
 NEWS  3  Feb 24  PCTGEN now available on STN
 NEWS  4  Feb 24  TEMA now available on STN
 NEWS  5  Feb 26  NTIS now allows simultaneous left and right truncation
 NEWS  6  Feb 26  PCTFULL now contains images
 NEWS  7  Mar 04  SDI PACKAGE for monthly delivery of multifile SDI results
 NEWS  8  Mar 24  PATDPAFULL now available on STN
 NEWS  9  Mar 24  Additional information for trade-named substances without
                  structures available in REGISTRY
 NEWS 10  Apr 11  Display formats in DGENE enhanced
 NEWS 11  Apr 14  MEDLINE Reload
 NEWS 12  Apr 17  Polymer searching in REGISTRY enhanced
 NEWS 13  Jun 13  Indexing from 1947 to 1956 added to records in CA/CAPLUS
 NEWS 14  Apr 21  New current-awareness alert (SDI) frequency in
                  WPIDS/WPINDEX/WPIX
 NEWS 15  Apr 28  RDISCLOSURE now available on STN
 NEWS 16  May 05  Pharmacokinetic information and systematic chemical names
                  added to PHAR
 NEWS 17  May 15  MEDLINE file segment of TOXCENTER reloaded
 NEWS 18  May 15  Supporter information for ENCOMPPAT and ENCOMPLIT updated
 NEWS 19  May 19  Simultaneous left and right truncation added to WSCA
 NEWS 20  May 19  RAPRA enhanced with new search field, simultaneous left and
                  right truncation
 NEWS 21  Jun 06  Simultaneous left and right truncation added to CBNB
 NEWS 22  Jun 06  PASCAL enhanced with additional data
 NEWS 23  Jun 20  2003 edition of the FSTA Thesaurus is now available
 NEWS 24  Jun 25  HSDB has been reloaded
 NEWS 25  Jul 16  Data from 1960-1976 added to RDISCLOSURE
 NEWS 26  Jul 21  Identification of STN records implemented
 NEWS 27  Jul 21  Polymer class term count added to REGISTRY
 NEWS 28  Jul 22  INPADOC: Basic index (/BI) enhanced; Simultaneous Left and
                  Right Truncation available
 NEWS 29  AUG 05  New pricing for EUROPATFULL and PCTFULL effective
                  August 1, 2003
 NEWS 30  AUG 13  Field Availability (/FA) field enhanced in BEILSTEIN
 NEWS 31  AUG 15  PATDPAFULL: one FREE connect hour, per account, in
                  September 2003
 NEWS 32  AUG 15  PCTGEN: one FREE connect hour, per account, in
                  September 2003
 NEWS 33  AUG 15  RDISCLOSURE: one FREE connect hour, per account, in
                  September 2003
 NEWS 34  AUG 15  TEMA: one FREE connect hour, per account, in
                  September 2003

 NEWS EXPRESS  April 4 CURRENT WINDOWS VERSION IS V6.01a, CURRENT
               MACINTOSH VERSION IS V6.0b(ENG) AND V6.0Jb(JP),
               AND CURRENT DISCOVER FILE IS DATED 01 APRIL 2003
```

```
 NEWS HOURS    STN Operating Hours Plus Help Desk Availability
 NEWS INTER    General Internet Information
 NEWS LOGIN    Welcome Banner and News Items
 NEWS PHONE    Direct Dial and Telecommunication Network Access to STN
 NEWS WWW      CAS World Wide Web Site (general information)

Enter NEWS followed by the item number or name to see news on that
specific topic.

  All use of STN is subject to the provisions of the STN Customer
  agreement.  Please note that this agreement limits use to scientific
  research.  Use for software development or design or implementation
  of commercial gateways or other similar uses is prohibited and may
  result in loss of user privileges and other penalties.

  * * * * * * * * * * * * * * STN  Columbus  * * * * * * * * * * * * * * * *

FILE 'HOME' ENTERED AT 09:50:49 ON 15 AUG 2003

=> file .biotech
COST IN U.S. DOLLARS                                   SINCE FILE     TOTAL
                                                            ENTRY   SESSION
FULL ESTIMATED COST                                          0.21      0.21

FILE 'MEDLINE' ENTERED AT 09:51:02 ON 15 AUG 2003

FILE 'BIOSIS' ENTERED AT 09:51:02 ON 15 AUG 2003
COPYRIGHT (C) 2003 BIOLOGICAL ABSTRACTS INC.(R)

FILE 'BIOTECHDS' ENTERED AT 09:51:02 ON 15 AUG 2003
COPYRIGHT (C) 2003 THOMSON DERWENT AND INSTITUTE FOR SCIENTIFIC INFORMATION

FILE 'CAPLUS' ENTERED AT 09:51:02 ON 15 AUG 2003
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
COPYRIGHT (C) 2003 AMERICAN CHEMICAL SOCIETY (ACS)

FILE 'EMBASE' ENTERED AT 09:51:02 ON 15 AUG 2003
COPYRIGHT (C) 2003 Elsevier Science B.V.  All rights reserved.

=> e blankenbecler richard/in
'IN' IS NOT A VALID EXPAND FIELD CODE FOR FILE 'MEDLINE'
'IN' IS NOT A VALID EXPAND FIELD CODE FOR FILE 'EMBASE'
E1             3      BLANKENBECKLER DAVID L/IN
E2             2      BLANKENBECKLER NICOLE LEE/IN
E3             5 --> BLANKENBECLER RICHARD/IN
E4             1      BLANKENBERG ANDREI/IN
E5             3      BLANKENBERG FRANCIS G/IN
E6             1      BLANKENBERG JUERGEN/IN
E7             3      BLANKENBORG STEPHANUS GERARDUS JOHANNES/IN
E8             2      BLANKENBURG BERND/IN
E9             2      BLANKENBURG CHARLES/IN
E10            1      BLANKENBURG GERLINDE/IN
E11            3      BLANKENBURG GUENTER/IN
E12            1      BLANKENBURG HANS ALWIN/IN
The indicated field code is not available for EXPAND in this
file.  To see a list of valid EXPAND field codes, enter HELP
SFIELDS at an arrow prompt (=>).

=> s e3
'IN' IS NOT A VALID FIELD CODE
'IN' IS NOT A VALID FIELD CODE
L1             5 "BLANKENBECLER RICHARD"/IN
```

```
=> d l1 1-5

L1   ANSWER 1 OF 5  BIOSIS  COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
AN   2001:179283  BIOSIS
DN   PREV200100179283
TI   Radial gradient contact lenses.
AU   Blankenbecler, Richard (1); Manhart, Paul K.
CS   (1) 974 Cottrell Way, Stanford, CA, 94305 USA
PI   US 6089711 July 18, 2000
SO   Official Gazette of the United States Patent and Trademark Office Patents,
     (July 18, 2000) Vol. 1236, No. 3, pp. No Pagination. e-file.
     ISSN: 0098-1133.
DT   Patent
LA   English

L1   ANSWER 2 OF 5  CAPLUS  COPYRIGHT 2003 ACS on STN
AN   2001:758693  CAPLUS
TI   A color separation optical plate for use with display and sensor panels
IN   Blankenbecler, Richard; Gourley, Helen; Levis, Maurice
PA   USA
SO   PCT Int. Appl.
     CODEN: PIXXD2
DT   Patent
LA   English
FAN.CNT 1
     PATENT NO.        KIND  DATE           APPLICATION NO.  DATE
     ---------------   ----  --------       ---------------  --------
PI   WO 2001077745      A1   20011018       WO 2000-US9686   20000411
         W:  AU, BR, CA, CN, JP, KR, MX, RU, US
         RW: AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL,
             PT, SE
PRAI WO 2000-US9686          20000411
RE.CNT  3       THERE ARE 3 CITED REFERENCES AVAILABLE FOR THIS RECORD
                ALL CITATIONS AVAILABLE IN THE RE FORMAT

L1   ANSWER 3 OF 5  CAPLUS  COPYRIGHT 2003 ACS on STN
AN   2001:748109  CAPLUS
DN   135:285367
TI   A method for protein structure alignment
IN   Blankenbecler, Richard; Ohlsson, Mattias; Peterson, Carsten;
     Ringner, Markus
PA   Board of Trustees of the Leland Stanford Junior University, USA
SO   PCT Int. Appl., 35 pp.
     CODEN: PIXXD2
DT   Patent
LA   English
FAN.CNT 1
     PATENT NO.        KIND  DATE           APPLICATION NO.  DATE
     ---------------   ----  --------       ---------------  --------
PI   WO 2001075436      A1   20011011       WO 2001-US10675  20010402
         W:  CA
         RW: AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL,
             PT, SE, TR
     US 2002111781      A1   20020815       US 2001-825441   20010402
     EP 1272840         A1   20030108       EP 2001-924605   20010402
         R:  AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT,
             IE, SI, LT, LV, FI, RO, MK, CY, AL, TR
PRAI US 2000-194203P    P    20000403
     WO 2001-US10675    W    20010402
RE.CNT  5       THERE ARE 5 CITED REFERENCES AVAILABLE FOR THIS RECORD
                ALL CITATIONS AVAILABLE IN THE RE FORMAT

L1   ANSWER 4 OF 5  CAPLUS  COPYRIGHT 2003 ACS on STN
AN   2000:560814  CAPLUS
```

THIS PAGE BLANK (USPTO)

```
TI   Color separation optical plate for use with lcd panels
IN   Blankenbecler, Richard; Levis, Maurice; Gourley, Helen
PA   USA
SO   U.S., 18 pp.
     CODEN: USXXAM
DT   Patent
LA   English
FAN.CNT 1
     PATENT NO.        KIND  DATE           APPLICATION NO.  DATE
     ---------------   ----  --------       ---------------  --------
PI   US 6104446         A    20000815       US 1996-769699   19961218
PRAI US 1996-769699          19961218
RE.CNT  19     THERE ARE 19 CITED REFERENCES AVAILABLE FOR THIS RECORD
               ALL CITATIONS AVAILABLE IN THE RE FORMAT

L1   ANSWER 5 OF 5  CAPLUS  COPYRIGHT 2003 ACS on STN
AN   1997:403287  CAPLUS
DN   127:22267
TI   Manufacture of refractive optical elements with radially graded properties
IN   Blankenbecler, Richard
PA   Blankenbecler, Richard, USA
SO   PCT Int. Appl., 34 pp.
     CODEN: PIXXD2
DT   Patent
LA   English
FAN.CNT 1
     PATENT NO.        KIND  DATE           APPLICATION NO.  DATE
     ---------------   ----  --------       ---------------  --------
PI   WO 9718170         A1   19970522       WO 1996-US18623  19961115
         W:  JP, KR
         RW: AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE
PRAI US 1995-558420          19951116


=> e ohlsson m
E1            1      OHLSONS/BI
E2           56      OHLSSON/BI
E3            0 -->  OHLSSON M/BI
E4            2      OHLSSONII/BI
E5            4      OHLSTADT/BI
E6            3      OHLSTEDT/BI
E7            3      OHLSTEIN/BI
E8            1      OHLSTORF/BI
E9            3      OHLU/BI
E10           1      OHLUCHNUTI/BI
E11           1      OHLUCHUTI/BI
E12           3      OHLUSENI/BI

=> e ohlsson mattias
E1            1      OHLSONS/BI
E2           56      OHLSSON/BI
E3            0 -->  OHLSSON MATTIAS/BI
E4            2      OHLSSONII/BI
E5            4      OHLSTADT/BI
E6            3      OHLSTEDT/BI
E7            3      OHLSTEIN/BI
E8            1      OHLSTORF/BI
E9            3      OHLU/BI
E10           1      OHLUCHNUTI/BI
E11           1      OHLUCHUTI/BI
E12           3      OHLUSENI/BI

=> e ohlsson m/in
'IN' IS NOT A VALID EXPAND FIELD CODE FOR FILE 'MEDLINE'
```

```
'IN' IS NOT A VALID EXPAND FIELD CODE FOR FILE 'EMBASE'
E1            1      OHLSSON LARS/IN
E2            4      OHLSSON LENNART G/IN
E3            1 --> OHLSSON M/IN
E4            1      OHLSSON MATTIAS/IN
E5            3      OHLSSON O A/IN
E6            1      OHLSSON O E/IN
E7            1      OHLSSON OLAF/IN
E8            1      OHLSSON OLOF A/IN
E9            1      OHLSSON OSCAR O/IN
E10           1      OHLSSON PER AAKE/IN
E11           1      OHLSSON PER AKE/IN
E12           1      OHLSSON S/IN
The indicated field code is not available for EXPAND in this
file.  To see a list of valid EXPAND field codes, enter HELP
SFIELDS at an arrow prompt (=>).

=> d e4
'E4' IS NOT A VALID FORMAT
In a multifile environment, a format can only be used if it is valid
in at least one of the files.  Refer to file specific help messages
or the STNGUIDE file for information on formats available in
individual files.
REENTER DISPLAY FORMAT FOR ALL FILES (FILEDEFAULT):
REENTER DISPLAY FORMAT FOR ALL FILES (FILEDEFAULT):filedefault

L1   ANSWER 1 OF 5  BIOSIS  COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
AN   2001:179283  BIOSIS
DN   PREV200100179283
TI   Radial gradient contact lenses.
AU   Blankenbecler, Richard (1); Manhart, Paul K.
CS   (1) 974 Cottrell Way, Stanford, CA, 94305 USA
PI   US 6089711 July 18, 2000
SO   Official Gazette of the United States Patent and Trademark Office Patents,
     (July 18, 2000) Vol. 1236, No. 3, pp. No Pagination. e-file.
     ISSN: 0098-1133.
DT   Patent
LA   English


=> s protein (s) align?
L2        21855 PROTEIN (S) ALIGN?

=> s (protein or ?peptide) (s) align?
LEFT TRUNCATION IGNORED FOR '?PEPTIDE' FOR FILE 'BIOTECHDS'
L3        24619 (PROTEIN OR ?PEPTIDE) (S) ALIGN?
Left truncation is not valid in the specified search field in the
specified file.  The term has been searched without left truncation.
Examples: '?TERPEN?' would be searched as 'TERPEN?' and '?FLAVONOID'
would be searched as 'FLAVONOID.'

If you are searching in a field that uses implied proximity, and you
used a truncation symbol after a punctuation mark, the system may
interpret the truncation symbol as being at the beginning of a term.
Implied proximity is used in search fields indexed as single words,
for example, the Basic Index.

=> s structure (a) l3
PROXIMITY OPERATOR LEVEL NOT CONSISTENT WITH
FIELD CODE - 'AND' OPERATOR ASSUMED 'STRUCTURE (A) L13'
PROXIMITY OPERATOR LEVEL NOT CONSISTENT WITH
FIELD CODE - 'AND' OPERATOR ASSUMED 'STRUCTURE (A) L14'
PROXIMITY OPERATOR LEVEL NOT CONSISTENT WITH
FIELD CODE - 'AND' OPERATOR ASSUMED 'STRUCTURE (A) L15'
```

```
PROXIMITY OPERATOR LEVEL NOT CONSISTENT WITH
FIELD CODE - 'AND' OPERATOR ASSUMED 'STRUCTURE (A) L16'
PROXIMITY OPERATOR LEVEL NOT CONSISTENT WITH
FIELD CODE - 'AND' OPERATOR ASSUMED 'STRUCTURE (A) L17'
L4       10381 STRUCTURE (A) L3

=> s structure (s) algin?
L5        1253 STRUCTURE (S) ALGIN?

=> s l4 and l5
L6           4 L4 AND L5

=> d l6 1-4

L6   ANSWER 1 OF 4  BIOSIS  COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
AN   2002:188924  BIOSIS
DN   PREV200200188924
TI   Identification of amino acid motifs important for epimerase activity of
     the Pseudomonas aeruginosa alginate modifying enzyme, AlgG.
AU   Douthit, S. A. (1); Franklin, M. J. (1)
CS   (1) Montana State University, Bozeman, MT USA
SO   Abstracts of the General Meeting of the American Society for Microbiology,
     (2001) Vol. 101, pp. 278. http://www.asmusa.org/mtgsrc/generalmeeting.htm.
     print.
     Meeting Info.: 101st General Meeting of the American Society for
     Microbiology Orlando, FL, USA May 20-24, 2001
     ISSN: 1060-2011.
DT   Conference
LA   English

L6   ANSWER 2 OF 4  BIOSIS  COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
AN   1987:171767  BIOSIS
DN   BA83:90208
TI   PURIFICATION AND STRUCTURAL PROPERTIES OF AN EXTRACELLULAR 1-4-BETA-D
     MANNURONAN-SPECIFIC ALGINATE LYASE FROM A MARINE BACTERIUM.
AU   ROMEO T; PRESTON J F III
CS   DEP. MICROBIOL. CELL SCI., UNIV. FLA., GAINESVILLE, FLA. 32611.
SO   BIOCHEMISTRY, (1986 (RECD 1987)) 25 (26), 8385-8391.
     CODEN: BICHAW. ISSN: 0006-2960.
FS   BA; OLD
LA   English

L6   ANSWER 3 OF 4  CAPLUS  COPYRIGHT 2003 ACS on STN
AN   2000:763565  CAPLUS
DN   134:53008
TI   Characterization of alginate lyase from Pseudomonas syringae pv. syringae
AU   Preston, Lori A.; Wong, T. Y.; Bender, Carol L.; Schiller, Neal L.
CS   Division of Biomedical Sciences, University of California, Riverside, CA,
     92521, USA
SO   Journal of Bacteriology (2000), 182(21), 6268-6271
     CODEN: JOBAAY; ISSN: 0021-9193
PB   American Society for Microbiology
DT   Journal
LA   English
RE.CNT  29    THERE ARE 29 CITED REFERENCES AVAILABLE FOR THIS RECORD
              ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6   ANSWER 4 OF 4  CAPLUS  COPYRIGHT 2003 ACS on STN
AN   1996:647091  CAPLUS
DN   125:294427
TI   Cloning, sequencing and overexpression in Escherichia coli of the alginate
     lyase-encoding aly gene of Pseudomonas alginovora: identification of three
     classes of alginate lyases
AU   Chavagnat, Frederic; Duez, Colette; Guinand, Micheline; Potin, Philippe;
```

```
     Barbeyron, Tristan; Henrissat, Bernard; Wallach, Jean; Ghuysen, Jean-Marie
CS   Lab. Biohimie Analytique, Univ. Claude Bernard, Villeurbanne, F-69622, Fr.
SO   Biochemical Journal (1996), 319(2), 575-583
     CODEN: BIJOAK; ISSN: 0264-6021
PB   Portland Press
DT   Journal
LA   English


=> d his

     (FILE 'HOME' ENTERED AT 09:50:49 ON 15 AUG 2003)

     FILE 'MEDLINE, BIOSIS, BIOTECHDS, CAPLUS, EMBASE' ENTERED AT 09:51:02 ON
     15 AUG 2003
                E BLANKENBECLER RICHARD/IN
L1            5 S E3
                E OHLSSON M
                E OHLSSON MATTIAS
                E OHLSSON M/IN
L2        21855 S PROTEIN (S) ALIGN?
L3        24619 S (PROTEIN OR ?PEPTIDE) (S) ALIGN?
L4        10381 S STRUCTURE (A) L3
L5         1253 S STRUCTURE (S) ALGIN?
L6            4 S L4 AND L5

=> s structure (s) align?
L7       18438 STRUCTURE (S) ALIGN?

=> s l4 and l7
L8        6639 L4 AND L7

=> s atomic distance
L9         481 ATOMIC DISTANCE

=> s stom? and distance?
L10       1596 STOM? AND DISTANCE?

=> s atom? distance?
L11       1740 ATOM? DISTANCE?

=> s l8 and l11
L12          4 L8 AND L11

=> d l12 1-4

L12  ANSWER 1 OF 4  BIOSIS  COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
AN   1992:499497  BIOSIS
DN   BA94:118022
TI   THREE-DIMENSIONAL STRUCTURE IN SOLUTION OF ACYL-COENZYME A
     BINDING PROTEIN FROM BOVINE LIVER.
AU   ANDERSEN K V; POULSEN F M
CS   CARLSBERG LAB., KEMISK AFDELING, GAMLE CARLSBERG VEJ 10, DK-2500 VALBY,
     COPENHAGEN, DEN.
SO   J MOL BIOL, (1992) 226 (4), 1131-1141.
     CODEN: JMOBAK. ISSN: 0022-2836.
FS   BA; OLD
LA   English

L12  ANSWER 2 OF 4  CAPLUS  COPYRIGHT 2003 ACS on STN
AN   1996:590946  CAPLUS
DN   125:269873
TI   Method and system for protein modeling
IN   Srinivasan, Subhashini; Sudarsanam, Padmanaban
```

THIS PAGE BLANK (USPTO)

```
PA   Immunex Corporation, USA
SO   U.S., 21 pp., Cont.-in-part of U.S. 5, 453, 397.
     CODEN: USXXAM
DT   Patent
LA   English
FAN.CNT 3
     PATENT NO.        KIND  DATE           APPLICATION NO.  DATE
     ---------------   ----  --------       ---------------  --------
PI   US 5557535         A    19960917       US 1994-234812   19940428
     US 5453937         A    19950926       US 1993-55050    19930428
     US 5884230         A    19990316       US 1996-663809   19960614
PRAI US 1993-55050           19930428
     US 1994-234812          19940428

L12  ANSWER 3 OF 4  CAPLUS  COPYRIGHT 2003 ACS on STN
AN   1995:420461  CAPLUS
DN   122:181416
TI   Computer method and system for protein modeling
IN   Srinivasan, Subhashini; Sudarsanam, Padmanaban
PA   Immunex Corp., USA
SO   PCT Int. Appl., 52 pp.
     CODEN: PIXXD2
DT   Patent
LA   English
FAN.CNT 3
     PATENT NO.        KIND  DATE           APPLICATION NO.  DATE
     ---------------   ----  --------       ---------------  --------
PI   WO 9425860         A1   19941110       WO 1994-US4822   19940428
         W:  AU, CA, FI, JP, KR, NO, NZ
         RW: AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE
     US 5453937         A    19950926       US 1993-55050    19930428
     AU 9467799         A1   19941121       AU 1994-67799    19940428
PRAI US 1993-55050           19930428
     WO 1994-US4822          19940428

L12  ANSWER 4 OF 4  EMBASE  COPYRIGHT 2003 ELSEVIER SCI. B.V. on STN
AN   92340912  EMBASE
DN   1992340912
TI   Three-dimensional structure in solution of acyl-coenzyme A
     binding protein from bovine liver.
AU   Andersen K.V.; Poulsen F.M.
CS   Carlsberg Laboratorium, Kemisk Afdeling, Gamle Carlsberg Vej 10,DK-2500
     Valby, Copenhagen, Denmark
SO   Journal of Molecular Biology, (1992) 226/4 (1131-1141).
     ISSN: 0022-2836  CODEN: JMOBAK
CY   United Kingdom
DT   Journal; Article
FS   029     Clinical Biochemistry
LA   English
SL   English


=> d his

     (FILE 'HOME' ENTERED AT 09:50:49 ON 15 AUG 2003)

     FILE 'MEDLINE, BIOSIS, BIOTECHDS, CAPLUS, EMBASE' ENTERED AT 09:51:02 ON
     15 AUG 2003
                E BLANKENBECLER RICHARD/IN
L1            5 S E3
                E OHLSSON M
                E OHLSSON MATTIAS
                E OHLSSON M/IN
L2        21855 S PROTEIN (S) ALIGN?
```

```
L3         24619 S (PROTEIN OR ?PEPTIDE) (S) ALIGN?
L4         10381 S STRUCTURE (A) L3
L5          1253 S STRUCTURE (S) ALGIN?
L6             4 S L4 AND L5
L7         18438 S STRUCTURE (S) ALIGN?
L8          6639 S L4 AND L7
L9           481 S ATOMIC DISTANCE
L10         1596 S STOM? AND DISTANCE?
L11         1740 S ATOM? DISTANCE?
L12            4 S L8 AND L11

=> s mean field
L13       20394 MEAN FIELD

=> s l4 and l13
L14          14 L4 AND L13

=> dup rem l14
PROCESSING COMPLETED FOR L14
L15            8 DUP REM L14 (6 DUPLICATES REMOVED)

=> d l15 ti 1-8

L15   ANSWER 1 OF 8  BIOTECHDS COPYRIGHT 2003 THOMSON DERWENT/ISI on STN
TI    Constructing a library of recombinant antibodies useful as source of
      antibody candidates for screening antigens comprises clustering variable
      regions of antibodies having known 3-dimensional structures into
      structural ensembles;
         single chain antibody, Fab, or Fv fragment library construction useful
         for antigen screening

L15  ANSWER 2 OF 8  CAPLUS  COPYRIGHT 2003 ACS on STN  DUPLICATE 1
TI   A novel approach to local reliability of sequence alignments

L15  ANSWER 3 OF 8  CAPLUS  COPYRIGHT 2003 ACS on STN
TI   A method for protein structure alignment

L15  ANSWER 4 OF 8  BIOSIS  COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
     DUPLICATE 2
TI   Identifying sequence-structure pairs undetected by sequence
     alignments.

L15  ANSWER 5 OF 8     MEDLINE on STN                    DUPLICATE 3
TI   Protein sequence-structure alignment based
     on site-alignment probabilities.

L15  ANSWER 6 OF 8  BIOSIS  COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
     DUPLICATE 4
TI   Model building by comparison at CASP3: Using expert knowledge and computer
     automation.

L15  ANSWER 7 OF 8  BIOSIS  COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
     DUPLICATE 5
TI   Model building by comparison: A combination of expert knowledge and
     computer automation.

L15  ANSWER 8 OF 8  EMBASE  COPYRIGHT 2003 ELSEVIER SCI. B.V. on STN
TI   Dynamics of an integral membrane peptide: A deuterium NMR relaxation study
     of gramicidin.


=> d ibib ab l15 1-8

L15   ANSWER 1 OF 8  BIOTECHDS COPYRIGHT 2003 THOMSON DERWENT/ISI on STN
```

THIS PAGE BLANK (USPTO)

ACCESSION NUMBER: 2003-07192 BIOTECHDS
TITLE: Constructing a library of recombinant antibodies useful as source of antibody candidates for screening antigens comprises clustering variable regions of antibodies having known 3-dimensional structures into structural ensembles; single chain antibody, Fab, or Fv fragment library construction useful for antigen screening
AUTHOR: LUO P
PATENT ASSIGNEE: ABMAXIS INC
PATENT INFO: WO 2002084277 24 Oct 2002
APPLICATION INFO: WO 2002-US12202 17 Apr 2002
PRIORITY INFO: US 2001-284407 17 Apr 2001; US 2001-284407 17 Apr 2001
DOCUMENT TYPE: Patent
LANGUAGE: English
OTHER SOURCE: WPI: 2003-093043 [08]

AB DERWENT ABSTRACT:
NOVELTY - Constructing a library of recombinant antibodies, comprising clustering variable regions of a collection of antibodies having known 3D structures into at least two families of structural ensembles, each comprising at least two different antibody sequences but with substantially identical main chain conformations, is new.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the following: (1) constructing a library of recombinant antibodies by: (a) clustering variable regions of a collection of antibodies having known 3D structures into at least two families of structural ensembles, each comprising at least two different antibody sequences but with substantially identical main chain conformations; (b) selecting a representative structural template from each family of structural ensemble; (c) profiling a tester polypeptide sequence onto the representative structural template within each family of structural ensemble; and (d) selecting the tester antibody sequence that is compatible to the structural constraints of the representative structural template; (2) constructing a library of recombinant antibodies based on a target structural template, by: (a) providing a target structural template of a variable region of one or more antibodies; (b) profiling a tester polypeptide sequence onto the target structural template; and (c) selecting the tester polypeptide sequence that is structurally compatible with the target structural template; and (3) constructing a library of recombinant antibodies by: (a) providing a target sequence of a heavy chain or light chain variable region of a target antibody; (b) **aligning** the target sequence with a tester polypeptide sequence; and (c) selecting the tester polypeptide sequence that has at least 15 % sequence homology with the target sequence.

BIOTECHNOLOGY - Preferred Method: In constructing a library of recombinant antibodies, the collection of antibodies includes antibodies or immunoglobulins collected in a **protein** database selected from the **protein** data bank of Brookhaven National Laboratory, genbank at the National Institute of Health, and Swiss-PROT **protein** sequence database. The collection of antibodies having known 3D structures include antibodies having resolved X-ray crystal structures, NMR structures or 3D structures based on structural modeling. The variable regions of the collection of antibodies are the full length heavy chain or light chain variable regions or specific portions of the heavy chain or light chain variable region consisting of complementary determining region (CDR) and/or framework region (FR), where the CDR is CDR 1, CDR2, or CDR3 of an antibody, and FR is FR1, FR2, FR3, or FR4 of an antibody. The clustering step includes clustering the collection of antibodies such that the root mean square of the main chain conformations of antibody sequences in each family of the structural ensemble is less than 4 Angstrom, preferably between 0.1-4.0 Angstrom, and that the Z-score of the main chain conformations of antibody sequences in each family of the structural ensemble is more than 2, 3 or 4, preferably 2-8. The clustering step is implemented by an algorithm selected from CE, Monte Carlo and 3D clustering algorithms. Profiling includes reverse

threading the tester polypeptide sequence onto the representative structural template within each family of structural ensemble, and is implemented by a multiple sequence **alignment** algorithm such as HMM algorithm or PSI-BLAST. The representative structural template is adopted by a CDR region, and the profiling step includes profiling the tester polypeptide sequence that is a variable region of a human or non-human antibody onto the representative structural template within each family of structural ensemble. The representative structural template is adopted by a FR region, and the profiling step includes profiling the tester polypeptide sequence that is a variable region of a human antibody onto the representative structural template within each family of structural ensemble. The tester polypeptide sequence is a variable region of human germline antibody sequence, a sequence or a segment sequence of an expressed **protein**, or a region of a human antibody. The tester polypeptide sequence is selected by using an energy scoring function consisting of electrostatic interactions, van der Waals interactions, electrostatic solvation energy, solvent-accessible surface solvation energy, or conformational entropy, or by using a scoring function incorporating a forcefield selected from Amber forcefield, Charmm forcefield, Discover cvff forcefields, ECEPP forcefields, GROMOS forcefields, OPLS forcefields, MMFF94 forcefield, Tripose forcefield, the MM3 forcefield, Dreiding forcefield, and UNRES forcefield, and other knowledge-based statistical forcefield (**mean field**) and **structure**-based thermodynamic potential functions. The method further comprises building an amino acid positional variant profile of the selected tester polypeptide sequences, filtering out the variants with occurrence frequency lower than 3, preferably lower than 5, and combining the variants remained to produce a combinatorial library of antibody sequences. After introducing the DNA segment encoding the selected tester polypeptide into cells of a host organism, expressing the DNA segment in the host cells such that a recombinant antibody containing the selected polypeptide sequence is produced in the cells of the host organism, and selecting the recombinant antibody that binds to a target antigen with affinity higher than 10 to the power 6/M. The recombinant antibody is a fully assembled antibody, a Fab fragment, an Fv fragment, or a single chain antibody. The host organism is selected from bacteria, yeast, plant, insect, and mammal, and the target antigen is a small molecule, proteins, **peptide**, nucleic acid or polycarbohydrate. The target sequence is an FR region of the target antibody, and **alignment** includes **aligning** the tester polypeptide sequence that is the sequence or segment sequence of a human antibody **protein** with the target sequence. The tested polypeptide sequence having at least 25 or 35 % sequence homology with the target sequence is selected.

USE - The method is useful for constructing a library of artificial antibodies in silico which provides a structurally diverse and yet functionally more relevant source of antibody candidates which can then be screened for binding a wide variety of target molecules, including small molecules, and biomacromolecules such as **protein**, **peptide** and nucleic acids. The libraries constructed are useful as source of antibody candidates for further screening for novel antibody with high affinity against a wide range of antigens and having no or minimum immunogenecity to human subjects treated with antibody therapeutics.

ADVANTAGE - The new method provides the following advantages of mapping the functional space of proteins using diversity of libraries that are designed by sampling the diversity in shape space rather than in sequence space: **protein-protein** interactions between ligand and receptor, antigen and antibody are conducted in well-defined conformation in space; simplicity in **structure** repertoire makes it easy to map the functional diversity based on variation in its 3D space and simple to cluster seemly complicated sequences pools into distinct families for library construction; provides a simple and viable approach to map its functional space; and **structure**-based

construction of sequence libraries makes it possible to apply various methods developed in structural biology to filter apparent complexity in sequence spaces based on structural or physical principles, in addition to the tools used in sequence analysis that are largely relied on the principles of evolution. (119 pages)

L15 ANSWER 2 OF 8 CAPLUS COPYRIGHT 2003 ACS on STN DUPLICATE 1

ACCESSION NUMBER:	2002:520873 CAPLUS
DOCUMENT NUMBER:	137:165760
TITLE:	A novel approach to local reliability of sequence alignments
AUTHOR(S):	Schlosshauer, Maximilian; Ohlsson, Mattias
CORPORATE SOURCE:	Complex Systems Division, Department of Theoretical Physics, University of Lund, Lund, S-223 62, Swed.
SOURCE:	Bioinformatics (2002), 18(6), 847-854 CODEN: BOINFP; ISSN: 1367-4803
PUBLISHER:	Oxford University Press
DOCUMENT TYPE:	Journal
LANGUAGE:	English

AB Motivation: The pairwise alignment of biol. sequences obtained from an algorithm will in general contain both correct and incorrect parts. Hence, to allow for a valid interpretation of the alignment, the local trustworthiness of the alignment has to be quantified. Results: We present a novel approach that attributes a reliability index to every pair of residues, including gapped regions, in the optimal **alignment** of two **protein** sequences. The method is based on a fuzzy recast of the dynamic programming algorithm for sequence alignment in terms of **mean field** annealing. An extensive evaluation with structural ref. alignments not only shows that the probability for a pair of residues to be correctly aligned grows consistently with increasing reliability index, but moreover demonstrates that the value of the reliability index can directly be translated into an est. of the probability for a correct alignment.

REFERENCE COUNT: 18 THERE ARE 18 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L15 ANSWER 3 OF 8 CAPLUS COPYRIGHT 2003 ACS on STN

ACCESSION NUMBER:	2001:748109 CAPLUS
DOCUMENT NUMBER:	135:285367
TITLE:	A method for **protein structure alignment**
INVENTOR(S):	Blankenbecler, Richard; Ohlsson, Mattias; Peterson, Carsten; Ringner, Markus
PATENT ASSIGNEE(S):	Board of Trustees of the Leland Stanford Junior University, USA
SOURCE:	PCT Int. Appl., 35 pp. CODEN: PIXXD2
DOCUMENT TYPE:	Patent
LANGUAGE:	English
FAMILY ACC. NUM. COUNT:	1

PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2001075436	A1	20011011	WO 2001-US10675	20010402
W: CA				
RW: AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR				
US 2002111781	A1	20020815	US 2001-825441	20010402
EP 1272840	A1	20030108	EP 2001-924605	20010402
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, SI, LT, LV, FI, RO, MK, CY, AL, TR				

PRIORITY APPLN. INFO.: US 2000-194203P P 20000403
WO 2001-US10675 W 20010402

AB This invention provides a method for **protein structure alignment**. More particularly, the present invention provides a method for identification, classification and prediction of protein structures. The present invention involves two key ingredients. First, an energy or cost function formulation of the problem simultaneously in terms of binary (Potts) assignment variables and real-valued at. coordinates. Second, a minimization of the energy or cost function by an iterative method, where in each iteration (1) a **mean field** method is employed for the assignment variables and (2) exact rotation and/or translation of at. coordinates is performed, weighted with the corresponding assignment variables.

REFERENCE COUNT: 5 THERE ARE 5 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L15 ANSWER 4 OF 8 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN DUPLICATE 2

ACCESSION NUMBER: 2000:438438 BIOSIS
DOCUMENT NUMBER: PREV200000438438
TITLE: Identifying sequence-**structure** pairs undetected by sequence alignments.
AUTHOR(S): Miyazawa, Sanzo (1); Jernigan, Robert L.
CORPORATE SOURCE: (1) Faculty of Technology, Gunma University, Kiryu, Gunma, 376 Japan
SOURCE: Protein Engineering, (July, 2000) Vol. 13, No. 7, pp. 459-475. print.
ISSN: 0269-2139.
DOCUMENT TYPE: Article
LANGUAGE: English
SUMMARY LANGUAGE: English

AB We examine how effectively simple potential functions previously developed can identify compatibilities between sequences and structures of proteins for database searches. The potential function consists of pairwise contact energies, repulsive packing potentials of residues for overly dense arrangement and short-range potentials for secondary structures, all of which were estimated from statistical preferences observed in known **protein** structures. Each potential energy term was modified to represent compatibilities between sequences and structures for globular proteins. Pairwise contact interactions in a sequence-**structure alignment** are evaluated in a **mean field** approximation on the basis of probabilities of site pairs to be **aligned**. Gap penalties are assumed to be proportional to the number of contacts at each residue position, and as a result gaps will be more frequently placed on **protein** surfaces than in cores. In addition to minimum energy **alignments**, we use probability **alignments** made by successively **aligning** site pairs in order by pairwise **alignment** probabilities. The results show that the present energy function and **alignment** method can detect well both folds compatible with a given sequence and, inversely, sequences compatible with a given fold, and yield mostly similar **alignments** for these two types of sequence and **structure** pairs. Probability **alignments** consisting of most reliable site pairs only can yield extremely small root mean square deviations, and including less reliable pairs increases the deviations. Also, it is observed that secondary **structure** potentials are usefully complementary to yield improved **alignments** with this method. Remarkably, by this method some individual sequence-**structure** pairs are detected having only 5-20% sequence identity.

L15 ANSWER 5 OF 8 MEDLINE on STN DUPLICATE 3

ACCESSION NUMBER: 2001646599 MEDLINE
DOCUMENT NUMBER: 21557041 PubMed ID: 11700595
TITLE: **Protein** sequence-**structure alignment** based on site-**alignment** probabilities.

AUTHOR: Miyazawa S
CORPORATE SOURCE: Faculty of Technology, Gunma University, Kiryu, Gunma 376, Japan.. miyazawa@smlab.sci.gunma-u.ac.jp
SOURCE: GENOME INFORMATICS SERIES, (2000) 11 141-50.
Journal code: 9717234. ISSN: 0919-9454.
PUB. COUNTRY: Japan
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 200112
ENTRY DATE: Entered STN: 20011112
Last Updated on STN: 20020124
Entered Medline: 20011231

AB A **protein** sequence-**structure alignment** method for database searches is examined on how effectively this method together with a simple scoring function previously developed can identify compatibilities between sequences and structures of proteins. The scoring function consists of pairwise contact energies, repulsive packing potentials of residues for overly dense arrangement and short-range potentials for secondary structures. Pairwise contact interactions in a sequence-**structure** alignment are evaluated in a **mean field** approximation on the basis of probabilities of site pairs to be aligned. Gap penalties are assumed to be proportional to the number of contacts at each residue position, and as a result gaps will be more frequently placed on protein surfaces than in cores. In addition to minimum energy alignments, we use probability alignments made by successively aligning site pairs in order by pairwise alignment probabilities. Results show that the present energy function and alignment method can detect well both folds compatible with a given sequence and, inversely, sequences compatible with a given fold. Probability alignments consisting of most reliable site pairs only can yield small root mean square deviations, and including less reliable pairs increases the deviations. Remarkably, by this method some individual sequence-**structure** pairs are detected having only 5-20% sequence identity.

L15 ANSWER 6 OF 8 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN DUPLICATE 4
ACCESSION NUMBER: 1999:496261 BIOSIS
DOCUMENT NUMBER: PREV199900496261
TITLE: Model building by comparison at CASP3: Using expert knowledge and computer automation.
AUTHOR(S): Bates, Paul A. (1); Sternberg, Michael J.E.
CORPORATE SOURCE: (1) Biomolecular Modelling Laboratory, Imperial Cancer Research Fund, 44 Lincoln's Inn Fields, London, WC2A 3PX UK
SOURCE: Proteins, (1999) Vol. 0, No. SUPPL. 3, pp. 47-54.
ISSN: 0887-3585.
DOCUMENT TYPE: Article
LANGUAGE: English
SUMMARY LANGUAGE: English

AB Ten models were constructed for the comparative modeling section of the Critical Assessment of Techniques for **Protein Structure** Prediction-3 (CASP3). Sequence identity between each target and the best possible parent(s) ranged between 12% and 64%. The modeling protocol is a mixture of automated computer algorithms with human intervention at certain critical stages. In particular, intervention is required to check sequence **alignments** and the selection of parameters for various computer programs. Seven of the targets were constructed from single-parent templates, and three were constructed from multiple parents. The reasons for such a high ratio of modeling from single parents only are discussed. Models constructed from multiple parents were found to be more accurate than models constructed from single parents only. A novel loop-modeling algorithm is presented that consists of fragment database searches, several fragment libraries, and **mean-field**

calculations on representative fragment candidates.

L15 ANSWER 7 OF 8 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
DUPLICATE 5

ACCESSION NUMBER: 1998:130564 BIOSIS
DOCUMENT NUMBER: PREV199800130564
TITLE: Model building by comparison: A combination of expert knowledge and computer automation.
AUTHOR(S): Bates, Paul A.; Jackson, Richard M.; Sternberg, Michael J. E. (1)
CORPORATE SOURCE: (1) Biomolecular Modelling Lab., Imperial Cancer Res. Fund, Lincoln's Inn Fields, P.O. Box 123, London WC2A 3PX UK
SOURCE: Proteins, (1997) Vol. 0, No. SUPPL. 1, pp. 59-67.
ISSN: 0887-3585.
DOCUMENT TYPE: Article
LANGUAGE: English

AB The CASP blind trials (Critical Assessment of techniques for **protein Structure** Prediction) assess the accuracy of **protein** prediction that includes evaluation of comparative model building of **protein** structures. Comparative models of four proteins (T0001, T0003, T0017, and T0028) for CASP2 (held during 1996) were constructed using computer algorithms combined with visual inspection. Essentially the main-chain modelling involves construction of the target **structure** from rigid-body segments of homologues and loop fragments extracted from homologous and nonredundant databases. Side-chains were initially constructed by inheritance from the parent or from a rotamer library. Side-chain conformations were then refined using a novel **mean field** approach that includes solvation. Comparison of the models with the subsequently released X-ray structures identified the successes and limitations of our approach. The most problematic area is the quality of the sequence **alignments** between parent(s) and target. In this respect the overinterpretation of the conserved features within homologous families can be misleading. Several features of our approach have a positive effect on the accuracy of the models. For T0003, inspection correctly identified that a lower sequence identity parent provides the best framework for this model. Loop selection worked well where a homologous **protein** fragment was used, but that the use of nonredundant fragment library remains problematic for hinge movements and displacements in secondary **structure** elements relative to the parent. Side-chain refinement improved residue conformations relative to the initial model. Use of limited energy minimization improved the stereochemical quality of the model without increasing the RMS deviation. This study has identified methods that are effective and areas requiring further attention to improve model building by comparison.

L15 ANSWER 8 OF 8 EMBASE COPYRIGHT 2003 ELSEVIER SCI. B.V. on STN

ACCESSION NUMBER: 94144117 EMBASE
DOCUMENT NUMBER: 1994144117
TITLE: Dynamics of an integral membrane peptide: A deuterium NMR relaxation study of gramicidin.
AUTHOR: Prosser R.S.; Davis J.H.
CORPORATE SOURCE: Department of Physics, University of Guelph,Guelph, Ont. N1G 2W1, Canada
SOURCE: Biophysical Journal, (1994) 66/5 (1429-1440).
ISSN: 0006-3495 CODEN: BIOJAU
COUNTRY: United States
DOCUMENT TYPE: Journal; Article
FILE SEGMENT: 029 Clinical Biochemistry
LANGUAGE: English
SUMMARY LANGUAGE: English

AB Solid state deuterium (2H) NMR inversion-recovery and Jeener-Broekaert relaxation experiments were performed on oriented multilamellar dispersions consisting of 1,2-dilauroyl-sn-glycero-3-phosphatidylcholine

and 2H exchange-labeled gramicidin D, at a lipid to **protein** molar ratio (L/P) of 15:1, in order to study the dynamics of the channel conformation of the **peptide** in a liquid crystalline phase. Our dynamic model for the whole body motions of the **peptide** includes diffusion of the **peptide** around its helix axis and a wobbling diffusion around a second axis perpendicular to the local bilayer normal in a simple Maier-Saupe **mean field** potential. This anisotropic diffusion is characterized by the correlation times, .tau.(R is parallel with) and .tau.(R is perpendicular to). **Aligning** the bilayer normal perpendicular to the magnetic field and graphing the relaxation rate, 1/T(1Z), as a function of (1 - S(N-2H)/2), where S(N-2H)/2 represents the orientational order parameter, we were able to estimate the correlation time, .tau.(R is parallel with), for rotational diffusion. Although in the quadrupolar splitting, which varies as (3 cos2 .theta.(D) - 1), has in general two possible solutions to .theta.(D) in the range 0 .ltoreq. .theta.(D) .ltoreq. 90.degree., the 1/T(1Z) vs. (1 - S(N-2H)/2) curve can be used to determine a single value of .theta.(D) in this range. Thus, the 1/T(1Z) vs. (1 - S(N-2H)/2) profile can be used both to define the axial diffusion rate and to remove potential structural ambiguities in the splittings. The T(1Z) anisotropy permits us to solve for the two correlation times (.tau.(R is parallel with) = 6.8 x 10-9 s and .tau.(R is perpendicular to) = 6 x 10-6 s). The simulated parameters were corroborated by a Jeener-Broekaert experiment where the bilayer normal was parallel to the principal magnetic field. At this orientation the ratio, J2(2.omega.0)/J1(.omega.0) was obtained in order to estimate the strength of the restoring potential in a model- independent fashion. This measurement yields the rms angle, <.theta.2>(1/2) (= 16 .+-. 2.degree. at 34.degree.C), formed by the **peptide** helix axis and the average bilayer normal.

```
=> FIL STNGUIDE
COST IN U.S. DOLLARS                               SINCE FILE      TOTAL
                                                        ENTRY    SESSION
FULL ESTIMATED COST                                    114.00     114.21

DISCOUNT AMOUNTS (FOR QUALIFYING ACCOUNTS)         SINCE FILE      TOTAL
                                                        ENTRY    SESSION
CA SUBSCRIBER PRICE                                     -1.30      -1.30

FILE 'STNGUIDE' ENTERED AT 10:25:18 ON 15 AUG 2003
USE IS SUBJECT TO THE TERMS OF YOUR CUSTOMER AGREEMENT
COPYRIGHT (C) 2003 AMERICAN CHEMICAL SOCIETY, JAPAN SCIENCE
AND TECHNOLOGY CORPORATION, AND FACHINFORMATIONSZENTRUM KARLSRUHE

FILE CONTAINS CURRENT INFORMATION.
LAST RELOADED: Aug  8, 2003 (20030808/UP).

=> d his

     (FILE 'HOME' ENTERED AT 09:50:49 ON 15 AUG 2003)

     FILE 'MEDLINE, BIOSIS, BIOTECHDS, CAPLUS, EMBASE' ENTERED AT 09:51:02 ON
     15 AUG 2003
                E BLANKENBECLER RICHARD/IN
L1            5 S E3
                E OHLSSON M
                E OHLSSON MATTIAS
                E OHLSSON M/IN
L2        21855 S PROTEIN (S) ALIGN?
L3        24619 S (PROTEIN OR ?PEPTIDE) (S) ALIGN?
L4        10381 S STRUCTURE (A) L3
L5         1253 S STRUCTURE (S) ALGIN?
L6            4 S L4 AND L5
```

```
L7         18438 S STRUCTURE (S) ALIGN?
L8          6639 S L4 AND L7
L9           481 S ATOMIC DISTANCE
L10         1596 S STOM? AND DISTANCE?
L11         1740 S ATOM? DISTANCE?
L12            4 S L8 AND L11
L13        20394 S MEAN FIELD
L14           14 S L4 AND L13
L15            8 DUP REM L14 (6 DUPLICATES REMOVED)

     FILE 'STNGUIDE' ENTERED AT 10:25:18 ON 15 AUG 2003

=> s (binary or potts) assign?
MISSING OPERATOR POTTS) ASSIGN?
The search profile that was entered contains terms or
nested terms that are not separated by a logical operator.

=> s (binary or potts) (s) assign
             0 BINARY
             0 POTTS
             0 ASSIGN
L16          0 (BINARY OR POTTS) (S) ASSIGN

=> s binary (s) assign?
             0 BINARY
            78 ASSIGN?
L17          0 BINARY (S) ASSIGN?

=> s binary assign?
             0 BINARY
            78 ASSIGN?
L18          0 BINARY ASSIGN?
                 (BINARY(W)ASSIGN?)

=> d his

     (FILE 'HOME' ENTERED AT 09:50:49 ON 15 AUG 2003)

     FILE 'MEDLINE, BIOSIS, BIOTECHDS, CAPLUS, EMBASE' ENTERED AT 09:51:02 ON
     15 AUG 2003
                E BLANKENBECLER RICHARD/IN
L1             5 S E3
                E OHLSSON M
                E OHLSSON MATTIAS
                E OHLSSON M/IN
L2         21855 S PROTEIN (S) ALIGN?
L3         24619 S (PROTEIN OR ?PEPTIDE) (S) ALIGN?
L4         10381 S STRUCTURE (A) L3
L5          1253 S STRUCTURE (S) ALGIN?
L6             4 S L4 AND L5
L7         18438 S STRUCTURE (S) ALIGN?
L8          6639 S L4 AND L7
L9           481 S ATOMIC DISTANCE
L10         1596 S STOM? AND DISTANCE?
L11         1740 S ATOM? DISTANCE?
L12            4 S L8 AND L11
L13        20394 S MEAN FIELD
L14           14 S L4 AND L13
L15            8 DUP REM L14 (6 DUPLICATES REMOVED)

     FILE 'STNGUIDE' ENTERED AT 10:25:18 ON 15 AUG 2003
L16            0 S (BINARY OR POTTS) (S) ASSIGN
L17            0 S BINARY (S) ASSIGN?
L18            0 S BINARY ASSIGN?
```

```
=> file .biotech
COST IN U.S. DOLLARS                                  SINCE FILE      TOTAL
                                                           ENTRY    SESSION
FULL ESTIMATED COST                                         0.72     114.93

DISCOUNT AMOUNTS (FOR QUALIFYING ACCOUNTS)            SINCE FILE      TOTAL
                                                           ENTRY    SESSION
CA SUBSCRIBER PRICE                                         0.00      -1.30

FILE 'MEDLINE' ENTERED AT 10:32:21 ON 15 AUG 2003

FILE 'BIOSIS' ENTERED AT 10:32:21 ON 15 AUG 2003
COPYRIGHT (C) 2003 BIOLOGICAL ABSTRACTS INC.(R)

FILE 'BIOTECHDS' ENTERED AT 10:32:21 ON 15 AUG 2003
COPYRIGHT (C) 2003 THOMSON DERWENT AND INSTITUTE FOR SCIENTIFIC INFORMATION

FILE 'CAPLUS' ENTERED AT 10:32:21 ON 15 AUG 2003
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
COPYRIGHT (C) 2003 AMERICAN CHEMICAL SOCIETY (ACS)

FILE 'EMBASE' ENTERED AT 10:32:21 ON 15 AUG 2003
COPYRIGHT (C) 2003 Elsevier Science B.V.  All rights reserved.

=> s (binary or potts) (s) assign?
L19          689 (BINARY OR POTTS) (S) ASSIGN?

=> s structure (s) align?
L20        18438 STRUCTURE (S) ALIGN?

=> s l19 and l20
L21            5 L19 AND L20

=> dup rem l21
PROCESSING COMPLETED FOR L21
L22             3 DUP REM L21 (2 DUPLICATES REMOVED)

=> d ibib ab l22 1-3

L22  ANSWER 1 OF 3  BIOSIS  COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
     DUPLICATE 1
ACCESSION NUMBER:     2003:211514  BIOSIS
DOCUMENT NUMBER:      PREV200300211514
TITLE:                Crystal structure of Pseudomonas fluorescens mannitol
                      2-dehydrogenase: Evidence for a very divergent long-chain
                      dehydrogenase family.
AUTHOR(S):            Kavanagh, Kathryn L.; Klimacek, Mario; Nidetzky, Bernd;
                      Wilson, David K. (1)
CORPORATE SOURCE:     (1) Section of Molecular and Cellular Biology, University
                      of California, One Shields Avenue, Davis, CA, 95616, USA:
                      dave@alanine.ucdavis.edu USA
SOURCE:               Chemico-Biological Interactions, (February 1 2003) Vol.
                      143-144, pp. 551-558. print.
                      ISSN: 0009-2797.
DOCUMENT TYPE:        Article
LANGUAGE:             English
AB   Mannitol 2-dehydrogenase from Pseudomonas fluorescens (pfMDH) is a
     secondary alcohol dehydrogenase that catalyzes the reversible
     NAD(P)-dependent oxidation of D-mannitol to D-fructose, D-arabinitol to
     D-xylulose, and D-sorbitol to L-sorbose. It is a member of the mostly
     prokaryotic family of long-chain mannitol dehydrogenases that so far
     includes 66 members. Unlike other alcohol and polyol dehydrogenases that
```

utilize metal cofactors or a conserved active-site tyrosine for catalysis, an invariant lysine is the general base. The crystal **structure** of pfMDH in a **binary** complex with NAD(H) and a ternary complex with NAD(H) and D-mannitol have been determined to 1.7 and 1.8 ANG resolution respectively. Comparison of secondary **structure assignment** to sequence **alignments** suggest the shortest members of this family, mannitol-1-phosphate 5-dehydrogenases, retain core elements but lack secondary structural components found on the surface of pfMDH. The elements predicted to be absent are distributed throughout the primary sequence, implying that a simple truncation or fusion did not occur. The closest structural neighbors are 6-phosphogluconate dehydrogenase, UDP-glucose dehydrogenase, N-(1-D-carboxyethyl)-L-norvaline dehydrogenase, and glycerol-3-phosphate dehydrogenase. Although sequence identity is only a barely recognizable 7-10%, conservation of secondary structural elements as well as homologous residues that are contributed to the active site indicates they may be related by divergent evolution.

L22 ANSWER 2 OF 3 CAPLUS COPYRIGHT 2003 ACS on STN

ACCESSION NUMBER: 2001:748109 CAPLUS
DOCUMENT NUMBER: 135:285367
TITLE: A method for protein **structure alignment**
INVENTOR(S): Blankenbecler, Richard; Ohlsson, Mattias; Peterson, Carsten; Ringner, Markus
PATENT ASSIGNEE(S): Board of Trustees of the Leland Stanford Junior University, USA
SOURCE: PCT Int. Appl., 35 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2001075436	A1	20011011	WO 2001-US10675	20010402
W: CA				
RW: AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR				
US 2002111781	A1	20020815	US 2001-825441	20010402
EP 1272840	A1	20030108	EP 2001-924605	20010402
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, SI, LT, LV, FI, RO, MK, CY, AL, TR				

PRIORITY APPLN. INFO.: US 2000-194203P P 20000403
WO 2001-US10675 W 20010402

AB This invention provides a method for protein **structure alignment**. More particularly, the present invention provides a method for identification, classification and prediction of protein structures. The present invention involves two key ingredients. First, an energy or cost function formulation of the problem simultaneously in terms of **binary (Potts) assignment** variables and real-valued at. coordinates. Second, a minimization of the energy or cost function by an iterative method, where in each iteration (1) a mean field method is employed for the assignment variables and (2) exact rotation and/or translation of at. coordinates is performed, weighted with the corresponding assignment variables.

REFERENCE COUNT: 5 THERE ARE 5 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L22 ANSWER 3 OF 3 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN DUPLICATE 2

ACCESSION NUMBER: 1991:363481 BIOSIS
DOCUMENT NUMBER: BA92:51706
TITLE: DIHYDROFOLATE REDUCTASE SEQUENTIAL RESONANCE ASSIGNMENTS

USING 2D AND 3D NMR AND SECONDARY STRUCTURE DETERMINATION IN SOLUTION.

AUTHOR(S): CARR M D; BIRDSALL B; FRENKIEL T A; BAUER C J; JIMENEZ-BARBERO J; POLSHAKOV V I; MCCORMICK J E; ROBERTS G C K; FEENEY J

CORPORATE SOURCE: LAB. MOL. STRUCTURE AND BIOMED. NMR CENT., NATL. INST. MED. RES., LONDON NW7 1AA, UK.

SOURCE: BIOCHEMISTRY, (1991) 30 (25), 6330-6341.
CODEN: BICHAW. ISSN: 0006-2960.

FILE SEGMENT: BA; OLD

LANGUAGE: English

AB Three-dimensional (3D) heteronuclear NMR techniques have been used to make sequential 1H and 15N resonance **assignments** for most of the residues of Lactobacillus casei ddihydrofolate reductase (DHFR), a monomeric protein of molecular mass 18 300 Da. A uniformly 15N-labeled sample of the protein was preapred and its complex with methotrexate (MTX) studied by 3D 15N/1H nuclear Overhauser-heteronuclear multiple quantum coherence (NOESY-HMQC), Hartmann-Hahn-heteronuclear multiple quantum coherence (HOHAHA-MHQC), and HMQC-NOESY-HMQC experiments. These experiments overcame most of the spectral overlap problems caused by chemical shift degeneracies in 2D spectra and allowed the 1H-1H through-space and through-bond connectivities to be identified unambiguously, leading to the resonance **assignments**. The novel HMQC-NOESY-HMQC experiment allows NOE cross peaks to be detected between NH protons even when their 1H chemical shifts are degenerate as long as the amide 15N chemical shifts are nondegenerate. The 3D experiments, in combination with conventional 2D NOESY, COSY, and HOHAHA experiments on unlabelled and selectively deuterated DHFR, provide backbone **assignments** for 146 of the 162 residues and side-chain **assignments** for 104 residues of the protein. Data from the NOE-based experiments and identification of the slowly exchanging amide protons provide detailed information about the secondary **structure** of the **binary** complex of the protein with methotrexate. Sequential NHi-NHi+1 NOEs define four regions with helical **structure**. Two of these regions, residues 44-49 and 79-89, correspond to within one amino acid to helices C and E in the crystal **structure** of the DHFR.cntdot.methotrexate.cntdot.NADPH complex [Bolin et al. (1982) J. Biol. Chem. 257, 13650-13662], while the NMR-determined helix formed by residues 26-35 is about one turn shorter at the N-terminus than helix B in the crystal **structure**, which spans residues 23-34. Similarly, the NMR-determined helical region comprising residues 102-110 is somewhat offset from the crystal strucure's helix F, which encompasses residues 97-107. Regions of .beta.-sheet **structure** were characterized in the **binary** complex by strong .alpha.CHi0NHi+1 NOEs and by slowly exchanging amide protons. In addition, several long-range NOEs were identified linking together these stretches to form a .beta.-sheet. These elements **align** perfectly with corresponding elemetns in the crystal **structure** of the DHFR.cntdot.methotrexate.cntdot.NADPH complex, which contains an eight-stranded .beta.-sheet, indicating that the main body of the .beta.-sheet is preserved in the **binary** complex in solution.

```
=> d his

     (FILE 'HOME' ENTERED AT 09:50:49 ON 15 AUG 2003)

     FILE 'MEDLINE, BIOSIS, BIOTECHDS, CAPLUS, EMBASE' ENTERED AT 09:51:02 ON
     15 AUG 2003
                E BLANKENBECLER RICHARD/IN
L1            5 S E3
                E OHLSSON M
                E OHLSSON MATTIAS
                E OHLSSON M/IN
```

```
L2         21855 S PROTEIN (S) ALIGN?
L3         24619 S (PROTEIN OR ?PEPTIDE) (S) ALIGN?
L4         10381 S STRUCTURE (A) L3
L5          1253 S STRUCTURE (S) ALGIN?
L6             4 S L4 AND L5
L7         18438 S STRUCTURE (S) ALIGN?
L8          6639 S L4 AND L7
L9           481 S ATOMIC DISTANCE
L10         1596 S STOM? AND DISTANCE?
L11         1740 S ATOM? DISTANCE?
L12            4 S L8 AND L11
L13        20394 S MEAN FIELD
L14           14 S L4 AND L13
L15            8 DUP REM L14 (6 DUPLICATES REMOVED)

     FILE 'STNGUIDE' ENTERED AT 10:25:18 ON 15 AUG 2003
L16            0 S (BINARY OR POTTS) (S) ASSIGN
L17            0 S BINARY (S) ASSIGN?
L18            0 S BINARY ASSIGN?

     FILE 'MEDLINE, BIOSIS, BIOTECHDS, CAPLUS, EMBASE' ENTERED AT 10:32:21 ON
     15 AUG 2003
L19          689 S (BINARY OR POTTS) (S) ASSIGN?
L20        18438 S STRUCTURE (S) ALIGN?
L21            5 S L19 AND L20
L22            3 DUP REM L21 (2 DUPLICATES REMOVED)

=> s (binary or potts) (a) assign?
L23           24 (BINARY OR POTTS) (A) ASSIGN?

=> dup rem l23
PROCESSING COMPLETED FOR L23
L24           12 DUP REM L23 (12 DUPLICATES REMOVED)

=> d ibib ab l24 1-12

L24  ANSWER 1 OF 12     MEDLINE on STN                    DUPLICATE 1
ACCESSION NUMBER:    2002132245     MEDLINE
DOCUMENT NUMBER:     21857028   PubMed ID: 11867791
TITLE:               Systemic lupus erythematosus: diagnostic application of
                     magnetization transfer ratio histograms in patients with
                     neuropsychiatric symptoms--initial results.
AUTHOR:              Dehmeshki Jamshid; Van Buchem Mark A; Bosma Gerlof P T;
                     Huizinga Tom W J; Tofts Paul S
CORPORATE SOURCE:    Institute of Neurology, University College London,
                     England.. j.dehmeshki@http-tech.com
SOURCE:              RADIOLOGY, (2002 Mar) 222 (3) 722-8.
                     Journal code: 0401260. ISSN: 0033-8419.
PUB. COUNTRY:        United States
DOCUMENT TYPE:       Journal; Article; (JOURNAL ARTICLE)
LANGUAGE:            English
FILE SEGMENT:        Abridged Index Medicus Journals; Priority Journals
ENTRY MONTH:         200205
ENTRY DATE:          Entered STN: 20020228
                     Last Updated on STN: 20020508
                     Entered Medline: 20020507
AB   PURPOSE: To explore the diagnostic potential of magnetization transfer
     ratio (MTR) histogram analysis in patients with neuropsychiatric systemic
     lupus erythematosus (SLE) by using multivariate discriminant analysis
     (MDA).  MATERIALS AND METHODS: Volumetric magnetization transfer imaging
     was performed in nine patients with active non-thromboembolic,
     neuropsychiatric SLE, 10 patients with SLE who had had neuropsychiatric
     SLE previously, 10 patients with SLE but no history of neuropsychiatric
     SLE, 10 patients with inactive multiple sclerosis, and 10 healthy control
```

subjects. For each subject, an MTR histogram of the whole brain was generated, and an MDA score was produced for each histogram. Each patient was assigned to a clinical subgroup on the basis of these MDA scores. For **assignment, binary** comparisons between subgroups were made. The accuracy of this classification method was assessed and compared with that of conventional MTR histogram analysis. RESULTS: With MDA, the success rate of binary classification was 60%-100%, depending on which two groups were compared. When the different clinical subgroups were separated, MDA parameters were always better than conventional MTR histogram parameters, with P values ranging from.05 to less than 1 x 10(-6) of those attained with the best conventional parameter. CONCLUSION: With MDA, MTR histograms of brain tissue may provide diagnostic information for individual patients in the clinical context of SLE.

```
L24  ANSWER 2 OF 12  CAPLUS  COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER:         2001:748109  CAPLUS
DOCUMENT NUMBER:          135:285367
TITLE:                    A method for protein structure alignment
INVENTOR(S):              Blankenbecler, Richard; Ohlsson, Mattias; Peterson,
                          Carsten; Ringner, Markus
PATENT ASSIGNEE(S):       Board of Trustees of the Leland Stanford Junior
                          University, USA
SOURCE:                   PCT Int. Appl., 35 pp.
                          CODEN: PIXXD2
DOCUMENT TYPE:            Patent
LANGUAGE:                 English
FAMILY ACC. NUM. COUNT:   1
PATENT INFORMATION:

     PATENT NO.        KIND  DATE          APPLICATION NO.  DATE
     ---------------   ----  --------      ---------------  --------
     WO 2001075436      A1   20011011      WO 2001-US10675  20010402
         W:  CA
         RW: AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL,
             PT, SE, TR
     US 2002111781      A1   20020815      US 2001-825441   20010402
     EP 1272840         A1   20030108      EP 2001-924605   20010402
         R:  AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT,
             IE, SI, LT, LV, FI, RO, MK, CY, AL, TR
PRIORITY APPLN. INFO.:                   US 2000-194203P  P  20000403
                                         WO 2001-US10675  W  20010402
```

AB This invention provides a method for protein structure alignment. More particularly, the present invention provides a method for identification, classification and prediction of protein structures. The present invention involves two key ingredients. First, an energy or cost function formulation of the problem simultaneously in terms of binary (**Potts) assignment** variables and real-valued at. coordinates. Second, a minimization of the energy or cost function by an iterative method, where in each iteration (1) a mean field method is employed for the assignment variables and (2) exact rotation and/or translation of at. coordinates is performed, weighted with the corresponding assignment variables.

REFERENCE COUNT: 5 THERE ARE 5 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

```
L24  ANSWER 3 OF 12  CAPLUS  COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER:         2000:202168  CAPLUS
DOCUMENT NUMBER:          133:101689
TITLE:                    Selecting protein targets for structural genomics of
                          Pyrobaculum aerophilum: validating automated fold
                          assignment methods by using binary hypothesis testing
AUTHOR(S):                Mallick, Parag; Goodwill, Kenneth E.; Fitz-Gibbon,
                          Sorel; Miller, Jeffrey H.; Eisenberg, David
```

THIS PAGE BLANK (USPTO)

CORPORATE SOURCE: UCLA-DOE Laboratory of Structural Biology and Molecular Medicine, Department of Chemistry and Biochemistry, Molecular Biology Institute, University of California, Los Angeles, CA, 90095-1570, USA
SOURCE: Proceedings of the National Academy of Sciences of the United States of America (2000), 97(6), 2450-2455
CODEN: PNASA6; ISSN: 0027-8424
PUBLISHER: National Academy of Sciences
DOCUMENT TYPE: Journal
LANGUAGE: English

AB Three-dimensional protein folds were assigned to all ORFs of the recently sequenced genome of the hyperthermophilic archaeon Pyrobaculum aerophilum. Binary hypothesis testing was used to est. a confidence level for each assignment. A sep. test was conducted to assign a probability for whether each sequence has a novel fold-i.e., one that is not yet represented in the exptl. database of known structures. Of the 2,130 predicted nontransmembrane proteins in this organism, 916 matched a fold at a cumulative 90% confidence level, and 245 could be assigned at a 99% confidence level. Likewise, 286 proteins were predicted to have a previously unobserved fold with a 90% confidence level, and 14 at a 99% confidence level. These statistically based tools are combined with homol. searches against the Online Mendelian inheritance in Man (OMIM) human genetics database and other protein databases for the selection of attractive targets for crystallog. or NMR structure detn.

REFERENCE COUNT: 29 THERE ARE 29 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L24 ANSWER 4 OF 12 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
ACCESSION NUMBER: 2000:8486 BIOSIS
DOCUMENT NUMBER: PREV200000008486
TITLE: Letter to the Editor: Backbone resonance assignments for the Fv fragment of the catalytic antibody NPN43C9 with bound p-nitrophenol.
AUTHOR(S): Kroon, Gerard J. A.; Martinez-Yamout, Maria A.; Krebs, Joseph F.; Chung, John; Dyson, H. Jane; Wright, Peter E. (1)
CORPORATE SOURCE: (1) Department of Molecular Biology, Skaggs Institute of Chemical Biology, Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA, 92037 USA
SOURCE: Journal of Biomolecular NMR, (Sept., 1999) Vol. 15, No. 1, pp. 83-84.
ISSN: 0925-2738.
DOCUMENT TYPE: Article; Letter
LANGUAGE: English

L24 ANSWER 5 OF 12 MEDLINE on STN DUPLICATE 2
ACCESSION NUMBER: 1998124212 MEDLINE
DOCUMENT NUMBER: 98124212 PubMed ID: 9464565
TITLE: **Binary assignments** of amino acids from pattern conservation.
AUTHOR: Irback A; Potthast F
CORPORATE SOURCE: Department of Theoretical Physics, University of Lund, Sweden.
SOURCE: PROTEIN ENGINEERING, (1997 Sep) 10 (9) 1013-7.
Journal code: 8801484. ISSN: 0269-2139.
PUB. COUNTRY: ENGLAND: United Kingdom
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 199804
ENTRY DATE: Entered STN: 19980422
Last Updated on STN: 19980422
Entered Medline: 19980414

AB We have developed a simple optimization procedure for **assigning**

binary values to amino acids. The binary values are determined by a maximization of the degree of pattern conservation in groups of closely related protein sequences. The maximization is carried out at fixed composition. For compositions approximately corresponding to an equipartition of the residues, the optimal encoding is found to be strongly correlated with hydrophobicity. The stability of the procedure is demonstrated. Our calculations are based upon sequences in the SWISS-PROT database.

L24 ANSWER 6 OF 12 MEDLINE on STN DUPLICATE 3

ACCESSION NUMBER:	93356672 MEDLINE
DOCUMENT NUMBER:	93356672 PubMed ID: 8352687
TITLE:	Prediction of visual function in eyes with mild to moderate posterior pole residua of retinopathy of prematurity. Cryotherapy for Retinopathy of Prematurity Cooperative Group.
AUTHOR:	Reynolds J; Dobson V; Quinn G E; Gilbert W S; Tung B; Robertson J; Flynn J T
CORPORATE SOURCE:	Department of Ophthalmology, Children's Hospital of Buffalo, NY.
CONTRACT NUMBER:	5 U10 EY05874 (NEI)
SOURCE:	ARCHIVES OF OPHTHALMOLOGY, (1993 Aug) 111 (8) 1050-6. Journal code: 7706534. ISSN: 0003-9950.
PUB. COUNTRY:	United States
DOCUMENT TYPE:	(CLINICAL TRIAL) Journal; Article; (JOURNAL ARTICLE) (MULTICENTER STUDY) (RANDOMIZED CONTROLLED TRIAL)
LANGUAGE:	English
FILE SEGMENT:	Abridged Index Medicus Journals; Priority Journals
ENTRY MONTH:	199309
ENTRY DATE:	Entered STN: 19931001 Last Updated on STN: 19931001 Entered Medline: 19930913

AB OBJECTIVE: The Multicenter Trial of Cryotherapy for Retinopathy of Prematurity (CRYO-ROP) assigned eyes with macular heterotopia to the "favorable" outcome category and eyes with retinal fold involving the macula to the "unfavorable" outcome category. This **binary assignment** did not agree well with measured visual acuity outcome. We tested the hypothesis that rating structural outcome on a continuum from less to more severe would improve prediction of visual acuity in eyes with macular heterotopia or retinal fold. DESIGN: Fundus photographs of the 69 eyes in the CRYO-ROP trial that had macular heterotopia (n = 55) or retinal fold (n = 14) at the 1-year follow-up were analyzed for severity of macular heterotopia, macular elevation, and pigmentary disturbances. Each physician author estimated each eye's predicted Snellen acuity, based on the photographic findings and clinical expertise. These results were compared with the grating acuity obtained at ages 1 and 3 1/2 years with the Teller Acuity Card procedure and with letter acuity obtained at age 3 1/2 years with the crowded HOTV test. PATIENTS: The 69 eyes were from 59 patients in the randomized portion of the CRYO-ROP trial. RESULTS: Although eyes with retinal fold tended to have greater visual impairment than eyes with macular heterotopia, there was a wide variation in acuity in both groups, and physicians were unable to predict visual acuity from retinal appearance. CONCLUSION: The physician cannot reliably predict either grating acuity or letter acuity in eyes with macular heterotopia or macular fold due to retinopathy of prematurity. There is no substitute for periodic visual acuity testing in these eyes.

L24 ANSWER 7 OF 12 MEDLINE on STN DUPLICATE 4

ACCESSION NUMBER:	92345531 MEDLINE
DOCUMENT NUMBER:	92345531 PubMed ID: 1637965
TITLE:	On the design of genome mapping experiments using short synthetic oligonucleotides.

AUTHOR: Fu Y X; Timberlake W E; Arnold J
CORPORATE SOURCE: Center for Demographic and Population Genetics, University of Texas, Houston 77225.
CONTRACT NUMBER: HG00337 (NHGRI)
SOURCE: BIOMETRICS, (1992 Jun) 48 (2) 337-59.
Journal code: 0370625. ISSN: 0006-341X.
PUB. COUNTRY: United States
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 199209
ENTRY DATE: Entered STN: 19920911
Last Updated on STN: 19980206
Entered Medline: 19920901

AB The DNA of an organism can be digested into smaller fragments, stored individually as clones in phage, for example, to create a clone library, and retrieved later, when needed. The original ordering of fragments is lost in the process of creating the library. Hence, it is important to be able to place clones in order according to their position along chromosome(s), and this process is referred to as "in vitro reconstruction" or "contig mapping" of an organismal genome. Clones in the phage library can be **assigned binary** call numbers by scoring each clone for hybridization (0 or 1) with a battery of short manufactured DNA sequences called synthetic oligonucleotides or with restriction enzyme digests of each clone. Those clones with similar call numbers are placed close together in the ordered library. We address the design question of how many clones and probes to use to carry out in vitro reconstruction of an organism's chromosomes. This physical mapping problem is placed in the context of coverage problems in geometrical probability. Various statistics are developed to summarize how an ordered library covers a chromosome, the extent of clone overlap, and the similarity between clone call numbers. Several tests for whether clones overlap are given, together with their power properties. A simulation study is used to determine how robust some of the tests for clone overlap are to model violations. Tables are presented for researchers to choose the number of clones and probes on the basis of both power and technical considerations surrounding the hybridization experiments.

L24 ANSWER 8 OF 12 MEDLINE on STN DUPLICATE 5
ACCESSION NUMBER: 91259328 MEDLINE
DOCUMENT NUMBER: 91259328 PubMed ID: 2045967
TITLE: Multiresolution, error-convergence halftone algorithm.
AUTHOR: Peli E
CORPORATE SOURCE: Physiological Optics Unit, Eye Research Institute, Boston, Massachusetts 02114.
CONTRACT NUMBER: EYR015957 (NEI)
SOURCE: JOURNAL OF THE OPTICAL SOCIETY OF AMERICA A. OPTICS AND IMAGE SCIENCE, (1991 Apr) 8 (4) 625-36.
Journal code: 8402086. ISSN: 0740-3232.
PUB. COUNTRY: United States
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 199107
ENTRY DATE: Entered STN: 19910802
Last Updated on STN: 19960129
Entered Medline: 19910717

AB A new halftone algorithm is described. The algorithm is designed for implementation on a parallel architecture in order to provide fast, progressive coding of moderate-resolution images. The design is based on a multiresolution, hierarchical, pyramidal structure. At each pyramid level, the binarized image is compared with the original, gray-tone image over a successively larger window of pixels for calculation of a weighted averaged error. Within each level, selected binarized pixels are tested

for possible changes in the **binary assignment**. The **binary assignment** is changed if the change results in a lower average error over the entire window. Varying the selection of test pixels can cause the same process to provide clustered-dot patterns and dithering. A comparison of performance with the best implementation of the error-propagation algorithm is presented visually. Quality is compared also in terms of isotropy of the texture and the appropriate blue-noise characteristics in areas of uniform gray tone. The benefits of this algorithm are realized with moderate-resolution display of the order of 512 dots X 512 dots. The processing can be carried out on smaller blocks since the results can be combined without any visible seams or edge effects.

L24 ANSWER 9 OF 12 CAPLUS COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER: 1990:555078 CAPLUS
DOCUMENT NUMBER: 113:155078
TITLE: Optimal design of multipurpose batch plants. 2. A decomposition solution strategy
AUTHOR(S): Papageorgaki, Savoula; Reklaitis, Gintaras V.
CORPORATE SOURCE: Sch. Chem. Eng., Purdue Univ., West Lafayette, IN, 47907, USA
SOURCE: Industrial & Engineering Chemistry Research (1990), 29(10), 2062-73
CODEN: IECRED; ISSN: 0888-5885
DOCUMENT TYPE: Journal
LANGUAGE: English
AB A mixed integer nonlinear programming NMINLPE formulation for the optimum design of a multipurpose plant is given in part 1. The complexity of the model makes the problem computationally intractable for direct soln. by existing MINLP soln. techniques. Consequently, a decompn. strategy is presented that alternately solves a MILP master problem, which dets. the values of the **binary assignment** variables for fixed campaign lengths, and a NLP subproblem, which performs equipment sizing and dets. the values of the campaign lengths. The effectiveness of the decompn. procedure is demonstrated with a no. of test problems that are solved in reasonable computation times.

L24 ANSWER 10 OF 12 CAPLUS COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER: 1985:443640 CAPLUS
DOCUMENT NUMBER: 103:43640
TITLE: Possible identification of the structural similarity of binary aqueous-nonaqueous mixtures according to the composition dependence of Walden products
AUTHOR(S): Kessler, Yu. M.; Bratishko, R. Kh.; Utkina, I. N.; Kumeev, R. S.
CORPORATE SOURCE: Inst. Khim. Nevodn., Rastvorov, USSR
SOURCE: Izvestiya Vysshikh Uchebnykh Zavedenii, Khimiya i Khimicheskaya Tekhnologiya (1985), 28(5), 118-20
CODEN: IVUKAR; ISSN: 0579-2991
DOCUMENT TYPE: Journal
LANGUAGE: Russian
AB A method is described for **assigning binary** aq. org. mixts. to structural groups on the characteristic concn. dependence of the Walden products. The method is applied to sulfolane, THF, DMF, DMSO, dioxane, tert-BuOH, and HMPA aq. solns. contg. <0.1 mol fraction org. solvent (using published cond. and viscosity data for LiCl, KCl, NaClO4, or LiNO3 solns.).

L24 ANSWER 11 OF 12 MEDLINE on STN DUPLICATE 6
ACCESSION NUMBER: 83176350 MEDLINE
DOCUMENT NUMBER: 83176350 PubMed ID: 6837941
TITLE: Tracheobronchial epithelium of the sheep: I. Quantitative light-microscopic study of epithelial cell abundance, and distribution.

AUTHOR: Mariassy A T; Plopper C G
SOURCE: ANATOMICAL RECORD, (1983 Mar) 205 (3) 263-75.
Journal code: 0370540. ISSN: 0003-276X.
PUB. COUNTRY: United States
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 198305
ENTRY DATE: Entered STN: 19900318
Last Updated on STN: 19900318
Entered Medline: 19830527

AB Glutaraldehyde-infused tracheas and airways of five castrated sheep were microdissected following the axial airway of the left cranial and caudal lobes. Airway branches were **assigned binary** numbers indicating their specific location in the tracheobronchial tree. Samples of known airway generation were resin embedded and examined by light-microscopy. Based on differences in cell morphology, staining properties, and distribution, eight major cell groups were recognized and quantified: four mucous cell categories (M1, M2, M3, and M4), ciliated, basal, Clara, and serous cells. The last cell category was restricted to submucosal glands. Tracheal epithelium had the most cells per unit length, primarily due to large numbers of basal cells. Basal cells are found in the epithelium of airways without cartilage or glands. The total mucous cell population (M1, M2, and M3) in proximal airways was relatively constant. M4 mucous cells were present in glands of proximal airways and in the epithelial lining of the airways without glands. The most distal airways were lined by Clara and ciliated cells. A small number of the most proximal noncartilaginous airways had mucous (M1, M2, M3, and M4), basal, and Clara cells sharing the epithelial lining. We conclude that in the sheep lung: (1) epithelial cell distribution does not correlate with airway wall components; (2) more than one type of secretory epithelial cell can share the lining of the same airway; and (3) Clara cell distribution is based on airway generation and proximity to alveoli.

L24 ANSWER 12 OF 12 CAPLUS
ACCESSION NUMBER: 1970:37338 CAPLUS
DOCUMENT NUMBER: 72:37338
TITLE: Vibrational spectra of N-chloroaziridine and N-bromoaziridine
AUTHOR(S): Russell, Joel Wolcott; Bishop, M.; Limburg, J.
CORPORATE SOURCE: Dep. of Chem., Oakland Univ., Rochester, MI, USA
SOURCE: Spectrochimica Acta, Part A: Molecular and Biomolecular Spectroscopy (1969), 25(12), 1929-39
CODEN: SAMCAS; ISSN: 1386-1425
DOCUMENT TYPE: Journal
LANGUAGE: English

AB The ir spectra of N-chloroaziridine, N-chloroaziridine-d4, and N-bromoazirdine were detd. along with the Raman spectra of all but N-chloroaziridine-d4. A complete vibrational assignment has been made for N-chloroaziridine; all but one fundamental assigned for N-bromoazirdine, and a partial assignment proposed for N-chloroaziridine-d4. Correlation of the vibrational frequencies of the haloaziridines with aziridine strongly supports the **Potts assignment** for azirdine with 1 minor revision.

=> d his

(FILE 'HOME' ENTERED AT 09:50:49 ON 15 AUG 2003)

FILE 'MEDLINE, BIOSIS, BIOTECHDS, CAPLUS, EMBASE' ENTERED AT 09:51:02 ON 15 AUG 2003

```
                E BLANKENBECLER RICHARD/IN
L1            5 S E3
```

```
                 E OHLSSON M
                 E OHLSSON MATTIAS
                 E OHLSSON M/IN
L2          21855 S PROTEIN (S) ALIGN?
L3          24619 S (PROTEIN OR ?PEPTIDE) (S) ALIGN?
L4          10381 S STRUCTURE (A) L3
L5           1253 S STRUCTURE (S) ALGIN?
L6              4 S L4 AND L5
L7          18438 S STRUCTURE (S) ALIGN?
L8           6639 S L4 AND L7
L9            481 S ATOMIC DISTANCE
L10          1596 S STOM? AND DISTANCE?
L11          1740 S ATOM? DISTANCE?
L12             4 S L8 AND L11
L13         20394 S MEAN FIELD
L14            14 S L4 AND L13
L15             8 DUP REM L14 (6 DUPLICATES REMOVED)

     FILE 'STNGUIDE' ENTERED AT 10:25:18 ON 15 AUG 2003
L16             0 S (BINARY OR POTTS) (S) ASSIGN
L17             0 S BINARY (S) ASSIGN?
L18             0 S BINARY ASSIGN?

     FILE 'MEDLINE, BIOSIS, BIOTECHDS, CAPLUS, EMBASE' ENTERED AT 10:32:21 ON
     15 AUG 2003
L19           689 S (BINARY OR POTTS) (S) ASSIGN?
L20         18438 S STRUCTURE (S) ALIGN?
L21             5 S L19 AND L20
L22             3 DUP REM L21 (2 DUPLICATES REMOVED)
L23            24 S (BINARY OR POTTS) (A) ASSIGN?
L24            12 DUP REM L23 (12 DUPLICATES REMOVED)

=> s potts (a) assign?
L25             2 POTTS (A) ASSIGN?

=> d ibib ab l25 1-2

L25  ANSWER 1 OF 2  CAPLUS  COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER:         2001:748109  CAPLUS
DOCUMENT NUMBER:          135:285367
TITLE:                    A method for protein structure alignment
INVENTOR(S):              Blankenbecler, Richard; Ohlsson, Mattias; Peterson,
                          Carsten; Ringner, Markus
PATENT ASSIGNEE(S):       Board of Trustees of the Leland Stanford Junior
                          University, USA
SOURCE:                   PCT Int. Appl., 35 pp.
                          CODEN: PIXXD2
DOCUMENT TYPE:            Patent
LANGUAGE:                 English
FAMILY ACC. NUM. COUNT:   1
PATENT INFORMATION:

     PATENT NO.        KIND  DATE          APPLICATION NO.  DATE
     ---------------   ----  --------      ---------------  --------
     WO 2001075436      A1   20011011      WO 2001-US10675  20010402
         W:  CA
         RW: AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL,
             PT, SE, TR
     US 2002111781      A1   20020815      US 2001-825441   20010402
     EP 1272840         A1   20030108      EP 2001-924605   20010402
         R:  AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT,
             IE, SI, LT, LV, FI, RO, MK, CY, AL, TR
PRIORITY APPLN. INFO.:                  US 2000-194203P  P  20000403
                                        WO 2001-US10675  W  20010402
```

AB This invention provides a method for protein structure alignment. More particularly, the present invention provides a method for identification, classification and prediction of protein structures. The present invention involves two key ingredients. First, an energy or cost function formulation of the problem simultaneously in terms of binary (**Potts) assignment** variables and real-valued at. coordinates. Second, a minimization of the energy or cost function by an iterative method, where in each iteration (1) a mean field method is employed for the assignment variables and (2) exact rotation and/or translation of at. coordinates is performed, weighted with the corresponding assignment variables.

REFERENCE COUNT: 5 THERE ARE 5 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L25 ANSWER 2 OF 2 CAPLUS COPYRIGHT 2003 ACS on STN

```
ACCESSION NUMBER:        1970:37338  CAPLUS
DOCUMENT NUMBER:         72:37338
TITLE:                   Vibrational spectra of N-chloroaziridine and
                         N-bromoaziridine
AUTHOR(S):               Russell, Joel Wolcott; Bishop, M.; Limburg, J.
CORPORATE SOURCE:        Dep. of Chem., Oakland Univ., Rochester, MI, USA
SOURCE:                  Spectrochimica Acta, Part A:  Molecular and
                         Biomolecular Spectroscopy (1969), 25(12), 1929-39
                         CODEN: SAMCAS; ISSN: 1386-1425
DOCUMENT TYPE:           Journal
LANGUAGE:                English
```

AB The ir spectra of N-chloroaziridine, N-chloroaziridine-d4, and N-bromoazirdine were detd. along with the Raman spectra of all but N-chloroaziridine-d4. A complete vibrational assignment has been made for N-chloroaziridine; all but one fundamental assigned for N-bromoazirdine, and a partial assignment proposed for N-chloroaziridine-d4. Correlation of the vibrational frequencies of the haloaziridines with aziridine strongly supports the **Potts assignment** for azirdine with 1 minor revision.

```
=> d his

     (FILE 'HOME' ENTERED AT 09:50:49 ON 15 AUG 2003)

     FILE 'MEDLINE, BIOSIS, BIOTECHDS, CAPLUS, EMBASE' ENTERED AT 09:51:02 ON
     15 AUG 2003
                E BLANKENBECLER RICHARD/IN
L1            5 S E3
                E OHLSSON M
                E OHLSSON MATTIAS
                E OHLSSON M/IN
L2        21855 S PROTEIN (S) ALIGN?
L3        24619 S (PROTEIN OR ?PEPTIDE) (S) ALIGN?
L4        10381 S STRUCTURE (A) L3
L5         1253 S STRUCTURE (S) ALGIN?
L6            4 S L4 AND L5
L7        18438 S STRUCTURE (S) ALIGN?
L8         6639 S L4 AND L7
L9          481 S ATOMIC DISTANCE
L10        1596 S STOM? AND DISTANCE?
L11        1740 S ATOM? DISTANCE?
L12           4 S L8 AND L11
L13       20394 S MEAN FIELD
L14          14 S L4 AND L13
L15           8 DUP REM L14 (6 DUPLICATES REMOVED)

     FILE 'STNGUIDE' ENTERED AT 10:25:18 ON 15 AUG 2003
L16           0 S (BINARY OR POTTS) (S) ASSIGN
```

```
L17           0 S BINARY (S) ASSIGN?
L18           0 S BINARY ASSIGN?

     FILE 'MEDLINE, BIOSIS, BIOTECHDS, CAPLUS, EMBASE' ENTERED AT 10:32:21 ON
     15 AUG 2003
L19         689 S (BINARY OR POTTS) (S) ASSIGN?
L20       18438 S STRUCTURE (S) ALIGN?
L21           5 S L19 AND L20
L22           3 DUP REM L21 (2 DUPLICATES REMOVED)
L23          24 S (BINARY OR POTTS) (A) ASSIGN?
L24          12 DUP REM L23 (12 DUPLICATES REMOVED)
L25           2 S POTTS (A) ASSIGN?

=> s energy (a) minimiz?
L26        9254 ENERGY (A) MINIMIZ?

=> s l4 and l26
L27         109 L4 AND L26

=> s l7 and l27
L28          83 L7 AND L27

=> dup rem l28
PROCESSING COMPLETED FOR L28
L29          44 DUP REM L28 (39 DUPLICATES REMOVED)

=> d ti l29 1-44

L29   ANSWER 1 OF 44  BIOTECHDS COPYRIGHT 2003 THOMSON DERWENT/ISI on STN
TI    Composition comprising isolated, recombinant Pseudomonas aeruginosa
      triosephosphate isomerase polypeptide useful for identifying modulators
      of the polypeptide which are used for treating bacteremia and keratitis;
         recombinant enzyme protein production via plasmid expression in host
         cell for use in disease therapy

L29  ANSWER 2 OF 44  CAPLUS  COPYRIGHT 2003 ACS on STN
TI   Mutagenesis of oxidoreductases for altering coenzyme-specificity and use
     for stereoselective synthesis

L29   ANSWER 3 OF 44  BIOTECHDS COPYRIGHT 2003 THOMSON DERWENT/ISI on STN
TI    Novel 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase protein or its
      functional protein subunit, in crystalline form, useful for identifying
      and designing inhibitors and activators of the protein;
         recombinant enzyme protein production and agonist and antagonist for
         use in disease therapy and drug screening

L29   ANSWER 4 OF 44  BIOTECHDS COPYRIGHT 2003 THOMSON DERWENT/ISI on STN
TI    New crystal of rifampicin bound to a core RNA polymerase (RNAP), useful
      for identifying an inhibitor of Taq RNAP activity that is used for
      preventing or treating a Taq RNAP mediated disease;
         RNA-polymerase structure determination for drug design and
         disease therapy

L29   ANSWER 5 OF 44  BIOTECHDS COPYRIGHT 2003 THOMSON DERWENT/ISI on STN
TI    Crystal of LuxS protein which is involved in production of autoinducer-2
      for identifying modulators useful for treating e.g. infection disease,
      stomach cancer, stomach ulcer and other intestinal complications;
         vector-mediated gene transfer, expression in host cell and computer
         bioinformatic software for recombinant protein production and drug
         screening

L29   ANSWER 6 OF 44  BIOTECHDS COPYRIGHT 2003 THOMSON DERWENT/ISI on STN
TI    Molecule or molecular complex useful for obtaining structural information
      of a molecule of unknown structure, comprises a portion of
```

THIS PAGE BLANK (USPTO)

Staphylococcus aureus FemA or FemA-like substrate binding surface/binding sites;
database, bioinformatic software and bioinformatic hardware for protein **structure** determination

L29 ANSWER 7 OF 44 CAPLUS COPYRIGHT 2003 ACS on STN
TI **Protein structure** prediction using a combination of sequence-based **alignment**, constrained **energy minimization**, and structural **alignment**

L29 ANSWER 8 OF 44 CAPLUS COPYRIGHT 2003 ACS on STN
TI Solution **structure** of a Kunitz-type chymotrypsin inhibitor isolated from the elapid snake Bungarus fasciatus

L29 ANSWER 9 OF 44 MEDLINE on STN DUPLICATE 1
TI Three-dimensional quantitative **structure**-activity relationship for several bioactive peptides searched by a convex hull-comparative molecular field analysis approach.

L29 ANSWER 10 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN DUPLICATE 2
TI Prediction of potential toxicity and side effect protein targets of a small molecule by a ligand-protein inverse docking approach.

L29 ANSWER 11 OF 44 MEDLINE on STN DUPLICATE 3
TI **Protein structure** prediction using a combination of sequence-based **alignment**, constrained **energy minimization**, and structural **alignment**.

L29 ANSWER 12 OF 44 EMBASE COPYRIGHT 2003 ELSEVIER SCI. B.V. on STN
TI **Protein structure** prediction using a combination of sequence-based **alignment**, constrained **energy minimization**, and structural **alignment**.

L29 ANSWER 13 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN DUPLICATE 4
TI The highly refined solution **structure** of the cytotoxic ribonuclease alpha-sarcin reveals the structural requirements for substrate recognition and ribonucleolytic activity.

L29 ANSWER 14 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN DUPLICATE 5
TI Homology-based molecular modelling of PLP-dependent histidine decarboxylase from Morganella morganii.

L29 ANSWER 15 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN DUPLICATE 6
TI Comparative molecular modeling on 3D-**structure** of opioid receptor-like 1 receptor.

L29 ANSWER 16 OF 44 MEDLINE on STN DUPLICATE 7
TI The third-dimensional **structure** of the complex between an Fv antibody fragment and an analogue of the main immunogenic region of the acetylcholine receptor: a combined two-dimensional NMR, homology, and molecular modeling approach.

L29 ANSWER 17 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN DUPLICATE 8
TI 3D modeling, ligand binding and activation studies of the cloned mouse delta, mu and kappa opioid receptors.

L29 ANSWER 18 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN DUPLICATE 9
TI Characterization of a type of beta-bend ribbon spiral generated by the

THIS PAGE BLANK (USPTO)

repeating (Xaa-Yaa-Aib-Pro) motif: The solution **structure** of harzianin HC IX, a 14-residue peptaibol forming voltage-dependent ion channels.

L29 ANSWER 19 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN DUPLICATE 10
TI **Structure** comparison of human glioma pathogenesis-related protein GliPr and the plant pathogenesis-related protein P14a indicates a functional link between the human immune system and a plant defense system.

L29 ANSWER 20 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN DUPLICATE 11
TI Homology model for the ligand-binding domain of the human estrogen receptor.

L29 ANSWER 21 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN DUPLICATE 12
TI Three-dimensional **structure** of toxin OSK1 from Orthochirus scrobiculosus scorpion venom.

L29 ANSWER 22 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN DUPLICATE 13
TI Comparative modeling and molecular dynamics studies of the delta, kappa and mu opioid receptors.

L29 ANSWER 23 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN DUPLICATE 14
TI Ligand-based **protein alignment** and isozyme specificity of glutathione S-transferase inhibitors.

L29 ANSWER 24 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN DUPLICATE 15
TI Model building by comparison: A combination of expert knowledge and computer automation.

L29 ANSWER 25 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN DUPLICATE 16
TI Modeling of the D1/D2 proteins and cofactors of the photosystem II reaction center: Implications for herbicide and bicarbonate binding.

L29 ANSWER 26 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN DUPLICATE 17
TI Thermodynamic prediction of conserved secondary **structure**: Application to the RRE element of HIV, the tRNA-like element of CMV and the mRNA of prion protein.

L29 ANSWER 27 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
TI Modeling spatial **structure** of obelin, a hydroid calcium-activated photoprotein.

L29 ANSWER 28 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN DUPLICATE 18
TI Prediction of the three-dimensional **structure** of the rap-1A protein from its homology to the ras-gene-encoded p21 protein.

L29 ANSWER 29 OF 44 EMBASE COPYRIGHT 2003 ELSEVIER SCI. B.V. on STN
TI Asymmetry in the **structure** of glycopeptide antibiotic dimers: NMR studies of the ristocetin A complex with a bacterial cell wall analogue.

L29 ANSWER 30 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN DUPLICATE 19
TI Predicting the **structure** of the light-harvesting complex II of

Rhodospirillum molischianum.

L29 ANSWER 31 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
DUPLICATE 20
TI Comparative modeling of the three-dimensional **structure** of Type
II antifreeze protein.

L29 ANSWER 32 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
DUPLICATE 21
TI A critical assessment of comparative molecular modeling of tertiary
structures of proteins.

L29 ANSWER 33 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
DUPLICATE 22
TI Purification, cloning, and sequencing of archaebacterial pyrophosphatase
from the extreme thermoacidophile Sulfolobus acidocaldarius.

L29 ANSWER 34 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
DUPLICATE 23
TI Comparative modelling of major house dust mite allergen Der p I:
Structure validation using an extended environmental amino acid
propensity table.

L29 ANSWER 35 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
DUPLICATE 24
TI Myelin P-0-glycoprotein: Predicted **structure** and interactions of
extracellular domain.

L29 ANSWER 36 OF 44 MEDLINE on STN DUPLICATE 25
TI Calculation of **protein** backbone geometry from alpha-carbon
coordinates based on **peptide**-group dipole **alignment**.

L29 ANSWER 37 OF 44 BIOTECHDS COPYRIGHT 2003 THOMSON DERWENT/ISI on STN
TI Molecular cloning of a hyperthermophilic archaebacterial
adenylate-kinase;
Sulfolobus acidocaldarius enzyme characterization (conference
abstract)

L29 ANSWER 38 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
DUPLICATE 26
TI ENERGY-OPTIMIZED **STRUCTURE** OF ANTIFREEZE PROTEIN AND ITS BINDING
MECHANISM.

L29 ANSWER 39 OF 44 MEDLINE on STN DUPLICATE 27
TI Comparative molecular modeling and crystallization of P-30 protein: a
novel antitumor protein of Rana pipiens oocytes and early embryos.

L29 ANSWER 40 OF 44 EMBASE COPYRIGHT 2003 ELSEVIER SCI. B.V. on STN
TI Molecular conformation of ubiquitinated structures and the implications
for regulatory function.

L29 ANSWER 41 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
DUPLICATE 28
TI THREE-DIMENSIONAL MODEL FOR STELLACYANIN A BLUE COPPER-PROTEIN.

L29 ANSWER 42 OF 44 MEDLINE on STN DUPLICATE 29
TI Protein **structure** prediction.

L29 ANSWER 43 OF 44 EMBASE COPYRIGHT 2003 ELSEVIER SCI. B.V. on STN
TI Understanding structural relationships in proteins of unsolved
three-dimensional **structure**.

L29 ANSWER 44 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
DUPLICATE 30

TI   A REFINED MODEL FOR THE VARIABLE DOMAINS FV OF THE J-539 BETA-1 6-D GALACTAN-BINDING IMMUNOGLOBULIN.

=> d ibib ab l29 1-20

L29   ANSWER 1 OF 44 BIOTECHDS COPYRIGHT 2003 THOMSON DERWENT/ISI on STN
ACCESSION NUMBER: 2003-18581 BIOTECHDS
TITLE: Composition comprising isolated, recombinant Pseudomonas aeruginosa triosephosphate isomerase polypeptide useful for identifying modulators of the polypeptide which are used for treating bacteremia and keratitis;
recombinant enzyme protein production via plasmid expression in host cell for use in disease therapy
AUTHOR: EDWARDS A; DHARAMSI A; VEDADI M; ALAM M Z; AWREY D; BEATTIE B; HOUSTON S; KIMBER M; LAM R; NETHERY K; LI Q; NECAKOV S; VALLEE F
PATENT ASSIGNEE: AFFINIUM PHARM INC
PATENT INFO: WO 2003027274 3 Apr 2003
APPLICATION INFO: WO 2002-CA1453 25 Sep 2002
PRIORITY INFO: US 2002-376917 1 May 2002; US 2001-324739 25 Sep 2001
DOCUMENT TYPE: Patent
LANGUAGE: English
OTHER SOURCE: WPI: 2003-468147 [44]
AB   DERWENT ABSTRACT:
NOVELTY - Composition (C1) comprising isolated, recombinant polypeptide (I) which comprises a 247 residue Pseudomonas aeruginosa triosephosphate isomerase polypeptide sequence (PS), given in the specification, an amino acid sequence having at least 95 % identity with PS, or an amino acid sequence encoded by a polynucleotide that hybridizes to complementary strand of a triosephosphate isomerase polynucleotide sequence (NS), is new.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for: (1) a sample comprising (I) labeled with a heavy atom or is enriched in at least one NMR isotope; (2) a crystallized (I), where the crystal has a p61 space group; (3) a crystallized complex comprising the crystallized (I) and a co-factor, where the complex is in crystal form; (4) a crystallized complex comprising the crystallized, (I) and a small organic molecule, where the complex is in crystal form; (5) a host cell comprising a nucleic acid encoding (I), where the host cell produces at least about 1 mg of the polypeptide per liter of culture and the polypeptide is at least about one-third soluble as measurable by gel electrophoresis; (6) an isolated recombinant polypeptide comprising 90 % identity to PS, or an amino acid sequence encoded by a polynucleotide that hybridizes under stringent conditions to the complementary strand of NS, where the polypeptide comprises one or more of the following amino acid residues at the specified position of the polypeptide: N at position 152, and S at position 233; (7) crystalline triosephosphate isomerase form P. aeruginosa comprising a hexagonal crystal having unit cell dimensions of a=77.349, c=175.528Angstrom, and space group p61, the unit cell containing two molecules per asymmetric unit; (8) a crystallized polypeptide comprising a **structure** of a polypeptide that is defined by a substantial portion of the atomic coordinates (AS1) of P. aeruginosa as given in the specification; (9) homology modeling a homolog of triosephosphate isomerase from P. aeruginosa, comprising: (a) **aligning** amino acid sequence of homolog of triosephosphate isomerase from P. aeruginosa with PS and incorporating the sequence of the homolog of triosephosphate isomerase from P. aeruginosa into a model of triosephosphate isomerase from P. aeruginosa derived from AS1 to yield a preliminary model of the homolog of triosephosphate isomerase from P. aeruginosa; (b) subjecting the preliminary model to **energy minimization** to yield an **energy minimized** model; and (c) remodeling regions of the **energy minimized** model where stereochemistry restraints are violated to

yield a final model of the homolog of triosephosphate isomerase from P. aeruginosa; (10) attempting to make a crystallized complex comprising a polypeptide and a modulator having a molecular weight of less than 5 kDa, involves crystallizing (I) so that crystals of the crystallized polypeptide will diffract x-rays to a resolution of 5Angstrom or better, and soaking the crystals in a solution comprising a potential modulator having molecular weight of less than 5 kDa; (11) incorporating a potential modulator in a crystal of a polypeptide, comprising placing a hexagonal crystal of triosephosphate isomerase from P.aeruginosa having unit cell dimensions of a=77.349, c=175.528Angstrom and space group p61 in a solution comprising the potential modulator; (12) a computer readable storage medium comprising digitally encoded structural data, where the data comprises AS1 for the backbone atoms of at least about six amino acid residues from a druggable region of triosephosphate isomerase from P. aeruginosa; (13) scalable three-dimensional configuration of points, at least portion of the points derived from some or all of AS1 for several amino acid residues from a druggable region of triosephosphate iosmerase from P. aeruginosa; (14) a scalable three-dimensional configuration of points comprising points having a root mean square deviation of less than 1.5Angstrom from AS1 for one or more of groups of atoms from a druggable region of triosephosphate isomerase from P. aeruginosa, as given in the specification; (15) a computer-assisted method for identifying an inhibitor of the activity of triosephosphate isomerase from P. aeruginosa, using the atomic coordinates AS1 identifying a potential modulator for the prevention or treatment of a P. aeruginosa related disease or disorder using the three dimensional **structure** of crystallized (I) preparing a potential modulator of a druggable region contained in a polypeptide using the atomic coordinates of PS apparatus for determining whether a compound is a potential modulator of a polypeptide; (16) making an inhibitor of triosephosphate isomerase activity; (17) a computer readable storage medium comprising digitally encoded data, that comprises structural coordinates for a druggable region that is structurally homologous to AS1 for a druggable region of triosephosphate isomerase from P. aeruginosa; and (18) a computer readable storage medium comprising digitally encoded data, where the data comprises a majority of the three-dimensional **structure** coordinates of AS1.

WIDER DISCLOSURE - (1) polynucleotide encoding (I), or its fragments;(2) a database comprising sequences of (I); (3) truncated (I); (4) polypeptide derived from (I); (5) generating sets of combinatorial mutants of (I); (6) modified (I); (7) isolated nucleic acids which differ from the polynucleotide encoding (I) due to degeneracy in the genetic code; (8) nucleic acids encoding proteins derived from P. aeruginosa and which have amino acid sequences evolutionarily related to (I); (9) expression vector comprising polynucleotide encoding (I); (10) nucleic acids encoding fusion proteins comprising (I); (11) transgenic non-human animal having which harbor a transgene comprising polynucleotide encoding (I); (12) computer readable medium comprising sequences of the nucleic acids, and database comprising nucleic acids sequences; (13) antibodies reactive with (I); (14) kits for detecting P. aeruginosa in biological sample; (15) vaccines comprising (I); and (16) array comprising polynucleotide encoding (I).

BIOTECHNOLOGY - Preferred Composition: (I) is at least 95 % pure as determined by gel electrophoresis. The polypeptide is purified to essential homogeneity. At least two-thirds of the polypeptide in the sample is soluble. The polypeptide is fused to at least one heterologous polypeptide that increases the solubility or stability of the polypeptide. The composition further comprises a matrix suitable for mass spectrometry. The matrix is a nicotinic acid derivative or a cinnamic acid derivative. **Protein** co-ordinate data is given in the patent specification.

ACTIVITY - Antibacterial; Antiinflammatory; Ophthalmological. No biological data is given.

MECHANISM OF ACTION - P. aeruginosa triosephosphate isomerase

polypeptide activity modulator.

USE - C1 is useful for identifying small molecules that bind to a polypeptide of C1. Crystallized (I) is useful for designing a modulator for prevention or treatment of P. aeruginosa related disease or disorder. The crystallized (I) is useful for obtaining structural information of the crystallized polypeptide, and for identifying a druggable region of a polypeptide. The three dimensional **structure** of crystallized (I) is useful for determining the crystal **structure** of a homolog of a polypeptide. The atomic coordinates of crystallized (I) is useful for obtaining structural information about a molecule or a molecular complex of unknown **structure**. (All claimed.) The inhibitors or modulators of triosephosphate isomerase from P. aeruginosa are useful for treating bacteremia, keratitis, osteomyelitis, otitis externa, conjunctivitis, endophthalmitis, alveolar necrosis, vascular invasion and burn infection.

ADMINISTRATION - The modulators are administered parenterally, by inhalation spray, topically, rectally, nasally, buccally, vaginally, or via and implanted reservoir. Dosage range from 0.01-100, preferably 0.5-75 mg/kg.(245 pages)

L29 ANSWER 2 OF 44 CAPLUS COPYRIGHT 2003 ACS on STN

ACCESSION NUMBER:	2003:42410 CAPLUS
DOCUMENT NUMBER:	138:102936
TITLE:	Mutagenesis of oxidoreductases for altering coenzyme-specificity and use for stereoselective synthesis
INVENTOR(S):	Nakai, Takahisa; Morikawa, Souichi; Kizaki, Noriyuki; Yasohara, Yoshihiko
PATENT ASSIGNEE(S):	Kaneka Corporation, Japan
SOURCE:	PCT Int. Appl., 59 pp. CODEN: PIXXD2
DOCUMENT TYPE:	Patent
LANGUAGE:	Japanese
FAMILY ACC. NUM. COUNT:	1
PATENT INFORMATION:	

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2003004653	A1	20030116	WO 2002-JP6688	20020702

W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM

RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG

PRIORITY APPLN. INFO.: JP 2001-200417 A 20010702
JP 2002-6303 A 20020115

AB A method for mutagenesis of enzymes, oxidoreductase in particular, to alter the coenzyme-specificity by genetic engineering, is disclosed. Novel carbonyl reductase variants capable of using NADH as coenzyme generated using the method, and their use in enzymic stereoselective synthesis of (S)-4-halo-3-hydroxybutyrate R1CH2C(:O)HC(R2)CO2R3 (I; R1=halo; R2=H; R3=(non)substituted alkyl or aryl) and optically active alcs. R1CH2CHOHC(R2)CO2R3 (R1, R2, R3 as in I) are provided. NADPH-dependent carbonyl reductase (S1) isolated from Candida magnoliae strain IFO0705 (CMCRD) catalyzing the redn. of Et 4-chloro-3-oxobutanoate (COBE) to Et (S)-4-chloro-3-hydroxybutanoate (CHBE), with a 100% enantiomeric excess, was genetically engineered to use NADH as coenzyme. Using multiple **protein** sequence **alignment**, comparative

THIS PAGE BLANK (USPTO)

mol. modeling, **energy minimization**, and mol. dynamic calcns., to carry out the three-dimensional **structure** estn., the residues potentially involved in coenzyme binding were identified. Site-directed mutagenesis was then performed to construct mutants having combinations of the following substitutions; S41A, S42A/R, S43Q/G/R, W63I/L/V/F/M, Y64D, N65I/V, S66N/L, Y47R, and A69E. All the mutants created used NADH as coenzyme to catalyze the redn. of Et 4-chloroacetoacetate, while no activity was retained when using NADH. Recombinant CMCRD expressed in E. coli converted Et 4-chloroacetoacetate to Et (S)-4-Chloro-3-hydroxybutyrate with 98 - 99% optical purity.

REFERENCE COUNT: 8 THERE ARE 8 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L29 ANSWER 3 OF 44 BIOTECHDS COPYRIGHT 2003 THOMSON DERWENT/ISI on STN

ACCESSION NUMBER: 2003-11657 BIOTECHDS

TITLE: Novel 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase protein or its functional protein subunit, in crystalline form, useful for identifying and designing inhibitors and activators of the protein;
recombinant enzyme protein production and agonist and antagonist for use in disease therapy and drug screening

AUTHOR: LOUIE G V; BUCHANAN S G; GAJIWALA K S; SAUDER M J

PATENT ASSIGNEE: STRUCTURAL GENOMIX INC

PATENT INFO: WO 2002102991 27 Dec 2002

APPLICATION INFO: WO 2002-US19451 17 Jun 2002

PRIORITY INFO: US 2001-299058 18 Jun 2001; US 2001-299058 18 Jun 2001

DOCUMENT TYPE: Patent

LANGUAGE: English

OTHER SOURCE: WPI: 2003-247867 [24]

AB DERWENT ABSTRACT:

NOVELTY - A 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MECPS) **protein** (I) or a functional MECPS **protein** subunit, in crystalline form, is new.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the following: (1) Producing (M1) a computer readable database; (2) A computer readable database (II) produced by (M1); (3) Producing (M2) a computer readable database; (4) A computer readable database (III) comprising a representation of a compound capable of binding a binding pocket of a MECPS **protein** or comprising a representation of a compound rationally designed to be capable of binding a binding pocket of a MECPS **protein**, produced using (II); (5) Producing (M3) a compound comprising a 3D molecular **structure** represented by the coordinates contained in a computer-readable database produced using (II); (6) Modulating MECPS **protein** activity by contacting the MECPS with a compound, where the compound is represented in a database producing using (II) or the compound is produced by (M3); (7) Identifying (M4) an activator or inhibitor of a **protein** that comprises a MECPS active site or binding pocket; (8) Producing an activator or inhibitor identified by (M4); (9) Producing (M5) a computer readable database comprising structural information about a molecule complex of unknown **structure**; (10) A computer readable database (IV) produced by (M5); (11) Electronic transmission of all or part of (II), (III) or (IV); (12) Homology modeling (M6) the **structure** of MECPS **protein** homolog; (13) Identifying (M7) a compound that binds MECPS **protein**; (14) Designing (M8) a compound that binds MECPS **protein**; (15) A machine-readable medium (V) embedded with information that corresponds to a 3D structural representation of (I), or embedded with the molecular structural coordinates given in the specification, or at least 50% or 80% of the coordinates, or with the molecular structural coordinates of a **protein** molecule comprising a MECPS **protein** binding pocket, where the binding pocket comprises at least three amino acids selected from Asp49, Asp59, Gly61, Ala103, Pro106, Lys107, Met108, Arg109, Thr135, Thr136, His37, Ser38, Ile60, Phe64, Asp66 and Leu79, having the structural coordinates

given in the specification, or by the structural coordinates of a binding pocket homolog, where the root mean square deviation of the backbone atoms of the amino acid residues of the binding pocket and the binding pocket homolog is less than 2.0Angstrom; (16) Electronically transmitting all or part of the information stored in (V); (17) Producing a mutant MECPS **protein**, having an altered property relative to a MECPS **protein**; and (18) Determining whether a compound binds MECPS **protein**.

WIDER DISCLOSURE - Also disclosed are: (1) Making (I); (2) Determining the 3D **structure** of (I); (3) Identifying or designing a modulator of MECPS activity; (4) Modulator of MECPS activity and its use; (5) Obtaining structural information about a molecule or molecular complex; and (6) Producing a co-crystal of a compound and MECPS.

BIOTECHNOLOGY - Preferred Crystal: (I) is a heavy-atom derivative crystal. (I) is a mutant. (I) is characterized by a set of structural coordinates that is substantially similar to the set of coordinates given in the specification. Preparation: (I) was prepared using standard recombinant techniques Preferred Methods: M1 comprises: (a) obtaining the three-dimensional (3D) molecular **structure** coordinates of a binding pocket of a MECPS **protein**, by obtaining 3D structural coordinates defining the **protein** or a binding pocket of the **protein**, from a crystal of the **protein**; and (b) introducing the structural coordinate into a computer to produce a database containing the molecular structural coordinates of the **protein** or binding pocket. M2 comprises: (a) generating a representation of binding pocket of a MECPS **protein** in a co-crystal with a compound, preferably a compound rationally designed to be capable of binding the binding pocket by preparing a binding test compound represented in a computer-readable database produced using (II); (b) forming a co-crystal of the compound with a **protein** comprising a binding pocket of a MECPS **protein**; (c) obtaining the structural coordinates of the binding pocket in the co-crystal; and (d) introducing the structural coordinates of the binding pocket or the co-crystal into a computer-readable database. The representation is selected from the compounds name, a chemical or molecular formula of the compound, a chemical **structure** of the compound, an identifier of the compound, and 3D molecular structural coordinates of the compound. Generating a 2D representation of the binding pocket comprises use of structural coordinates having a root mean square deviation of the backbone atoms of the amino acid residues of the binding pocket of less than 2.0Angstrom from the structural coordinates of the corresponding residues, given in the specification. At least one binding test compound is selected from: (i) selecting a compound from a small molecule database; (ii) modifying a known inhibitor, substrate, reaction intermediate or reaction product, or a portion of MECPS; (iii) assembling chemical fragments or groups into a compound, and (iv) de novo ligand design of the compound. Assessing if a test compound model fits comprises docking the model to the representation of the MECPS binding pocket and/or performing **energy minimization**. M3 comprises synthesizing the compound, where the compound fits a binding pocket of MECPS **protein**. M4 comprises: (a) producing a compound by M3; (b) contacting the compound with a **protein** that comprises MECPS active site or binding pocket; and (c) determining whether the potential modulator activates or inhibits the activity of the **protein**. M5 comprises: (a) generating an X-ray diffraction pattern from a crystallized form of the molecule or molecular complex, using a molecular replacement method to interpret the **structure** of the molecule, where the molecular replacement method uses the structural coordinates given in the specification, or its subset comprising a binding pocket, where the structural coordinates of the binding pocket are given in the specification, or structural coordinates having a root mean square derivation for the alpha-C atoms of the structural coordinates of less than 2.0 Angstrom; and (b) storing the coordinates of the resulting

structure in a computer-readable database. M6 comprises; (a) **aligning** the amino acid sequence of a MECPS **protein** homolog with an amino acid sequence of MECPS **protein**; (b) incorporating the sequence of the MECPS **protein** homolog into a model of the **structure** of MECPS **protein**, where the model has the same structural coordinates as the structural coordinates given in the specification, or where the structural coordinates of the model's alpha-C atoms have a root mean square deviation from the structural coordinates given in the specification, of less than 2.0Angstrom to yield to preliminary model of the homolog; (c) subjecting the preliminary model to **energy minimization** to yield an **energy minimized** model; and (d) remodeling regions of the **energy minimized** model where stereoelectrochemistry restraints are violated to yield a final model of the homolog. M7 comprises: (a) providing a computer modeling program with a set of structural coordinates or a 3D conformation for a molecule that comprises a binding pocket of MECPS **protein**, or its homolog providing the computer modeling program with a set of structural coordinates of a chemical entity; (b) using the computer modeling program to evaluate the potential binding or interfering interactions between the chemical entity and the binding pocket; and (c) determining whether the chemical entity potentially binds to or interferes with the **protein** or homolog. M7 further comprises: (a) computationally modifying the structural coordinates or 3D conformation of the chemical entity to improve the likelihood of binding to the binding pocket; and (b) determining if the modified chemical entity binds to or interferes with the **protein** or homolog. Determining if the chemical entity potentially binds to the molecule comprises performing a fitting operation between the chemical entity and a binding pocket of the **protein** or homolog, and computationally analyzing the results of the fitting operation to quantify the association between, or the interference with, the chemical entity and the binding pocket. A library of structural coordinates of chemical entities is used to identify a compound that binds. M8 comprises: (a) providing a computer modeling program with a set of structural coordinates, or a 3D confirmation derived from it, for a molecule that comprises a binding pocket comprising the structural coordinates of a binding pocket of MECPS **protein**, or its homolog; (b) computationally building a chemical entity represented by set of structural coordinates, and (c) determining whether the chemical entity is expected to bind to the molecule.

ACTIVITY - Antimicrobial. No supporting data provided.

MECHANISM OF ACTION - 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MECPS) Agonist/Antagonist.

USE - (II) is useful for producing a computer readable database comprising a representation of a compound capable of binding a binding pocket of a MECPS **protein** or comprising a representation of a compound rationally designed to be capable of binding a binding pocket of a MECPS **protein**. The methods are useful for producing a compound comprising a 3D molecular **structure** represented by the coordinates contained in a computer-readable database, modulating MECPS **protein** activity by contacting the MECPS with a compound, identifying an activator or inhibitor of a **protein** that comprises a MECPS active site or binding pocket, producing a mutant MECPS **protein**, having an altered property relative to a MECPS **protein**, and determining whether a compound binds MECPS **protein** (all claimed). (I) is useful for identifying and designing inhibitors and activators of MECPS, for designing anti-microbials that target the active site or a binding format of MECPS, or otherwise interfere with MECPS activity, or another activity in an associated biochemical, metabolic or anabolic pathway, or for rational drug design to identify and/or design compounds that binds MECPS for developing new therapeutic agents.

EXAMPLE - Preparation of crystals of 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MECPS) was as follows. An open-reading

frame for MECPS was amplified from Haemophilus influenzae genomic DNA by polymerase chain reaction (PCR) using the following primers: ATCAGAATTGGACACGGCTTTG and CTTGTCGGATTAAAAGAGCAACG. The PCR product (474 base pairs expected) was electrophoresed on a 1% agarose gel in TBE buffer and appropriate size band was excised from the gel and eluted. The eluted DNA was ligated for 5 minutes at room temperature with topoisomerase into pSB3-TOPO. The vector pSB3-TOPO was a topoisomerase-activated, modified version of pET26b. The resulting sequence of the gene after being ligated into the vector, from the Shine-Dalgarno sequence through the stop site and the original BamHI, site was as follows: AAGGAGGAGAT ATACATATGTCCCTT(ORF)AAGGGGGATCCCACCACCAC CACCACCACTGAGATCC. The MECPS expressed using this vector had three amino acids added to its N-terminal end (MSL) and 10 amino acids added to its C-terminal end (EGGSHHHHHH). A coding sequence for MECPs was amplified from H.influenzae genomic DNA by PCR reaction using the following primers: ATATATATCATATGTCCCTTATCAGAATTGGACACGGCTTTG and TATAGGATCCCCCTTCTTGTCGGATTAAAAGAGCAACG. The PCR product was digested with NdeI and BamHI, electrophoresed on a 1% agarose gel in TBE buffer and the appropriate size band was excised from the gel and eluted. The eluted DNA was ligated overnight with T4 DNA ligase at 16degreesC into pSB3, previously digested with NdeI and BamHI. The vector pSB3 was a modified version of pET26b. The resulting sequence of the gene after being ligated into the vector, from the Shine-Dalgarno sequence through the stop site and the original BamHI, site was as follows: AAGGAGGAGATATACATATGTCCCTT(ORF)AAGGGGGATCCCACCACCACCACCACCACTGAGATCC. Plasmids containing ligated inserts were transformed into chemically competent TOP10 cells. Colonies were then screened for inserts in the correct orientation and small DNA amounts were purified using a miniprep procedure from 2 ml cultures. The miniprep DNA was transformed into BL21(DE3) cells and plated onto petri dishes containing Luria Bertani (LB) agar with 30 microg/ml of kanamycin. Isolated, single, colonies were grown to mid-log phase and stored at -80degreesC in LB containing 15% glycerol. MECPS containing selenomethionine was over expressed in Escherichia coli by the addition of 200 microl 1M isopropyl-beta-D-thiogalactopyranoside (IPTG) per 500 ml culture of minimal broth plus selenomethionine, and the cultures were allowed to ferment overnight. The MECPS was purified. For crystals of H.influenzae, MECPS from which the molecular **structure** coordinates were obtained, it was found that a hanging drop containing 1-2 microl of MECPS polypeptide 15 mg/ml in 10 mM Hepes pH 7.5, 150 mM, 150 mM NaCl 1 mM betaME 10 mM methionine, 10% glycerol and an equal volume reservoir solution: 30% (v/v) MPD, 200 mM CaCl, and 100 mM sodium acetate, pH 4.5 in a sealed container containing 500 microL, reservoir solution, incubated for 2-5 days at 4-12degreesC provided diffraction quality crystals. (370 pages)

L29 ANSWER 4 OF 44 BIOTECHDS COPYRIGHT 2003 THOMSON DERWENT/ISI on STN
ACCESSION NUMBER: 2003-04198 BIOTECHDS
TITLE: New crystal of rifampicin bound to a core RNA polymerase (RNAP), useful for identifying an inhibitor of Taq RNAP activity that is used for preventing or treating a Taq RNAP mediated disease;
RNA-polymerase **structure** determination for drug design and disease therapy
AUTHOR: DARST S; CAMPBELL E
PATENT ASSIGNEE: UNIV ROCKEFELLER
PATENT INFO: WO 2002073192 19 Sep 2002
APPLICATION INFO: WO 2002-US7535 11 Mar 2002
PRIORITY INFO: US 2001-802755 9 Mar 2001; US 2001-802755 9 Mar 2001
DOCUMENT TYPE: Patent
LANGUAGE: English
OTHER SOURCE: WPI: 2002-723386 [78]
AB DERWENT ABSTRACT:
NOVELTY - A crystal of rifampicin bound to a core RNA polymerase (Rif-RNAP) that effectively diffracts X-rays for the determination of the

atomic coordinates to a resolution of better than 3.5 Angstroms, is new.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the following: (1) a scalable three-dimensional configuration of points, where a portion of the points are derived from **structure** coordinates of a portion of a Taq RNAP molecule or molecular complex or its homolog comprising a substrate binding pocket; (2) a machine-readable storage medium comprising a data storage material encoded with machine readable data which, when using a machine programmed with instructions for using the data, is capable of displaying a graphical three-dimensional representation of a molecule or molecular complex; (3) obtaining structural information about a molecule or a molecular complex of unknown **structure**; (4) homology modeling a Taq RNAP homolog; (5) a computer-assisted method for identifying or designing an inhibitor of RNAP activity; (6) a computer-assisted method for designing a modulator of RNAP activity de novo; (7) making a modulator or inhibitor of RNAP activity; (8) an inhibitor or modulator of RNAP activity; (9) a pharmaceutical composition comprising an inhibitor or modulator of RNAP activity or its salt and a carrier; (10) identifying an agent for inhibiting bacterial RNA polymerase or that inhibits bacterial growth, or for use as a modulator of bacterial RNA polymerase; (11) crystallizing a RNAP complex or its subunit or portion with a binding partner; (12) obtaining a crystal of an inhibitor bound to a core bacterial RNA polymerase; (13) identifying a compound that is predicted to inhibit bacterial RNA polymerase or bacterial growth; or (14) a computer having within its memory a representation of rifampicin bound to the core RNA polymerase or its portion of the Rif-RNAP molecular complex, comprising: (i) a machine-readable data storage medium comprising a data storage material encoded with machine-readable data, where the data comprises a portion of the structural coordinates, given in the specification;(ii) a working memory for storing instructions for processing the machine-readable data; (iii) a central processing unit coupled to the working memory and to the machine-readable data storage medium for processing the machine readable data into a three-dimensional representation of the Rif-RNAP molecular complex or its portion; and (iv) a display coupled to the central-processing unit for displaying the three-dimensional representation.

BIOTECHNOLOGY - Preferred Crystal: The crystal further comprises an omega subunit. It has a space group of P41212 and a unit cell of dimensions of a = b = 201 or c = 294 Angstrom. The core RNA polymerase is a thermophilic bacterial core RNA polymerase, particularly Thermus aquaticus. It comprises a beta' subunit, beta subunit or a pair of alpha subunits. Preferred Configuration: The scalable three-dimensional configuration of points are displayed as a holographic image, a stereodiagram, a model or a computer-displayed image. Preferred Data: The machine-readable data is capable of displaying a graphical three-dimensional representation of a molecule or molecular complex consisting of: (i) a molecule or molecular complex comprising a portion of a substrate binding pocket having the amino acids, given in the specification;(ii) a homolog to a Taq RNAP molecule or molecular complex, where the Taq RNAP molecule or molecular complex is represented by a portion of the **structure** coordinates, given in the specification. The substrate binding pocket defined by sets of points has a root mean square deviation of less than about 1.1 Angstrom or 1.5 Angstrom from points representing the backbone atoms of the amino acids as represented by **structure** coordinates or the side chain atoms and the Calpha atoms of the amino acids as represented by **structure** coordinates, given in the specifications. When combined with a second set of machine readable data, using a machine programmed with instructions for using the first set of data and the second set of data, the machine readable data can determine a portion of the **structure** coordinates corresponding to the second set of machine readable data, where the first set of data comprises a Fourier transform of a portion of the **structure** coordinates for Taq RNAP, given in the specification and the second set of data comprises an x-ray

diffraction pattern of a molecule or molecular complex of unknown **structure**. All of the points in the scalable three-dimensional configuration of points are derived from **structure** coordinates of a Taq RNAP molecule or molecular complex or from the backbone atoms of amino acids, given in the specification. A portion of points are derived from **structure** coordinates representing the locations of the backbone atoms of amino acids defining the binding pocket, comprising the amino acids, given in the specification, or of the side chain atoms and the Calpha atom of the amino acids defining a substrate binding pockets comprising the amino acids or their majority, given in the specification. Preferred Method: Obtaining structural information about a molecule or a molecular complex of unknown **structure** comprises: (a) crystallizing the molecule or molecular complex; (b) generating an x-ray diffraction pattern from the crystallized molecule or molecular complex; and (c) applying a portion of the **structure** coordinates set to the x-ray diffraction pattern to generate a three-dimensional electron density map of a portion of the molecule or molecular complex whose **structure** is unknown. Homology modeling a Taq RNAP homolog comprises: (a) **aligning** the amino acid sequence of a Taq RNAP homolog with an amino acid sequence of Taq RNAP and incorporating the sequence of the RNAP homolog into a model of Taq RNAP derived from **structure** coordinates given in the specification to yield a preliminary model of the Taq RNAP homolog; (b) subjecting the preliminary model to **energy minimization** to yield an **energy minimized** model; and (c) remodeling regions of the **energy minimized** model where stereochemistry restraints are violated to yield a final model of the Taq RNAP homolog. The computer-assisted method for identifying an inhibitor of RNAP activity comprises: (a) supplying a computer modeling application with a set of **structure** coordinates of a molecule or molecular complex, the molecule or molecular complex comprising a substrate binding pocket; and (b) supplying the computer modeling application with a set of **structure** coordinates of a chemical entity, and determining whether the chemical entity is expected to modulate the molecule or molecular complex, where modulation of the molecule or molecular complex is indicative of potential modulation of RNAP activity. Determining whether the chemical entity is a modulator expected to modulate the molecule or molecular complex comprises: (a) performing a fitting operation between the chemical entity and a binding pocket of the molecule or molecular complex; (b) computationally analyzing the results of the fitting operation to quantify the association between the chemical entity and the binding pocket; and (c) screening a library of chemical entities. The computer-assisted method for designing an inhibitor of RNAP activity comprises: (a) supplying a computer modeling application with the structural coordinates for two-thirds of the amino acids of a substrate binding pocket; (b) supplying the computer modeling application with a set of **structure** coordinates for a chemical entity; (c) evaluating the potential binding interactions between the chemical entity and substrate binding pocket of the molecule or molecular complex; (d) structurally modifying the chemical entity to yield a set of **structure** coordinates for a modified chemical entity; and (e) determining whether the modified chemical entity is an inhibitor expected to bind to or interfere with the molecule or molecular complex, where binding to or interfering with the molecule or molecular complex is indicative of potential inhibition of RNAP activity. The computer-assisted method for designing a modulator of RNAP activity de novo comprises: (a) supplying a computer modeling application with a set of **structure** coordinates of a molecule or molecular complex, the molecule or molecular complex comprising a substrate binding pocket, with up to three conservative amino acid substitutions of the amino acids; (b) computationally building a chemical entity represented by set of **structure** coordinates; and (c) determining whether the chemical entity is expected to modulate the molecule or molecular complex, where modulation of the molecule or molecular complex is

indicative of potential modulation of RNAP activity. The methods further comprises: (a) supplying or synthesizing the potential inhibitor; and (b) assaying the potential inhibitor to determine whether it inhibits RNAP activity. Making a modulator of RNAP activity comprises: (a) synthesizing a chemical entity to yield a modulator of KNAP activity, the chemical entity having been identified during a computer assisted process comprising supplying a computer modeling application with a set of **structure** coordinates of a molecule or molecular complex, the molecule or molecular complex comprising a portion of a Taq RNAP or RNAP-like substrate binding pocket; (b) supplying the computer modeling application with a set of **structure** coordinates of a chemical entity; and (c) determining whether the chemical entity is expected to modulate the molecule or molecular complex at a binding pocket. Making an inhibitor of RNAP activity comprises: (a) preparing a chemical entity to yield an inhibitor of RNAP activity, the chemical entity having been designed during a computer assisted process comprising supplying a computer modeling application with a set of **structure** coordinates of a molecule or molecular complex, the molecule or molecular complex comprising a portion of a Taq RNAP or RNAP-like substrate binding pocket; (b) supplying the computer modeling application with a set of **structure** coordinates for a chemical entity; (c) evaluating the potential binding interactions between the chemical entity and a binding pocket of the molecule or molecular complex; (d) structurally modifying the chemical entity to yield a set of **structure** coordinates for a modified chemical entity; and (e) determining whether the chemical entity is expected to bind to or interfere with the molecule or molecular complex at the binding pocket, where binding to or interfering with the molecule or molecular complex is indicative of potential inhibition of RNAP activity. Identifying an agent for inhibiting bacterial RNA polymerase or that inhibits bacterial growth, or for use as a modulator of bacterial RNA polymerase comprises: (a) obtaining a set of atomic coordinates defining the three-dimensional **structure** of rifampicin bound to the core RNA polymerase; (b) selecting a potential agent by performing rational drug design with a portion of the atomic coordinates obtained in step (a), where the selecting is performed in conjunction with computer modeling; (c) contacting the potential agent with a bacterial RNA polymerase or bacterial culture; (d) measuring the activity of the bacterial RNA polymerase or the growth of the bacterial culture in the absence or presence of the agent, where a potential agent is identified as an agent that inhibits bacterial RNA polymerase or the growth of bacterial culture when there is a decrease in the activity of the bacterial RNA polymerase in the presence of the agent relative to in its absence; (e) preparing a supplemental crystal containing the core RNA polymerase formed in the presence of the potential agent, where the crystal effectively diffracts X-rays for the determination of the atomic coordinates to a resolution of better than 5.0 Angstrom; (f) determining the three-dimensional coordinates of the supplemental crystal; (g) selecting a second generation agent by performing rational drug design with the three-dimensional coordinates determined for the supplemental crystal, where the selecting is performed in conjunction with computer modeling; (h) contacting the second generation agent with a eukaryotic RNA polymerase; (i) measuring the activity of the eukaryotic RNA polymerase, where an agent is identified as an agent for use as an inhibitor of bacterial RNA polymerase or of bacterial growth when there is no change in the activity of the eukaryotic RNA polymerase or in the proliferation of the eukaryotic cell in the presence of the agent, relative to its absence. Crystallizing a RNAP complex or its subunit or portion with a binding partner comprises: (a) providing purified RNAP at a concentration of about 1 mg/ml to about 50 mg/ml; (b) mixing the purified RNAP with a solution comprising saturated (NH4)2 to obtain a mixture; and (c) incubating the mixture as a hanging drop over the same solution. Obtaining a crystal of an inhibitor bound to a core bacterial RNA polymerase comprises: (a) growing the core bacterial RNA polymerase crystaI in a buffered solution containing 40 - 45 % saturated ammonium

sulfate, where a crystal forms; and (b) soaking the crystal in 2 M (NH4)2SO4, with the inhibitor, where a crystal of the inhibitor bound to the core bacterial RNA polymerase is formed. The inhibitor is rifampicin. The growing is performed by a method consisting of batch crystallization, vapor diffusion or microdialysis. Identifying a compound that is predicted to inhibit bacterial RNA polymerase or bacterial growth comprises: (a) defining the **structure** of rifampicin bound to the core RNA polymerase or a portion of the Rif-RNAP molecular complex by the atomic coordinates, given in the specification, where the portion of the molecular complex comprises sufficient structural information to perform step (b); (b) identifying a compound that is predicted to inhibit bacterial RNA polymerase or bacterial growth, where the identifying is performed using the **structure** defined in step (a); (c) contacting the compound with a bacterial RNA polymerase or with a bacterial culture; (d) measuring the activity of the bacterial RNA polymerase or the growth of the bacterial culture in the absence of the compound; (e) contacting the compound with a eukaryotic RNA polymerase or cell; and (f) measuring the activity of the eukaryotic RNA polymerase or the amount of proliferation of the eukaryotic cell. The specification contains 3-D **protein** structural data.

ACTIVITY - Antibacterial. No biological data is given.

MECHANISM OF ACTION - Taq RNAP-Inhibitor.

USE - The crystal is used in obtaining structural information about a molecule or molecular complex of unknown **structure**. New methods are used for: (i) homology modeling a Taq RNAP homolog; (ii) identifying and designing an inhibitor of RNAP activity; (iii) designing a modulator of RNAP activity de novo; (iv) making a modulator or inhibitor of RNAP activity; (v) identifying an agent for inhibiting bacterial RNA polymerase or that inhibits bacterial growth, or for use as a modulator of bacterial RNA polymerase; (vi) crystallizing a RNAP complex or its subunit or portion with a binding partner; (vii) obtaining a crystal of an inhibitor bound to a core bacterial RNA polymerase; and (viii) identifying a compound that is predicted to inhibit bacterial RNA polymerase or bacterial growth (all claimed). A composition comprising an inhibitor of Taq RNAP activity is useful for the prevention and treatment of Taq RNAP mediated disease.

ADMINISTRATION - Administration can be oral, parenteral, inhalation, topical, rectal, nasally, buccal, vaginal or via an implanted reservoir. Dose is 0.01 - 100 mg/kg body weight , preferably 0.5 - 75 mg/kg body weight per day.

EXAMPLE - Native Taq core DNA dependent RNA polymerase (RNAP) was purified and crystallized by standard methods. Crystals were subsequently soaked in stabilization solution with 0.1 mM rifampicin for 12 hours. Crystals were then prepared for cry-crystallography by soaking in stabilization solution containing 50 % (w/v) sucrose for 30 minutes before flash freezing in liquid nitrogen. Diffraction data was collected at the APS beamline SBC 19ID using 0.3 oscillations, and processed using DENZO (RTM) and SCALEPAK (RTM). The Taq core RNAP:Rif crystals were isomorphous with the native Taq core RNAP crystals. Strong electron density was observed in difference Fourier maps for the rifampicin which occupied a shallow pocket between beta structural domains 3 and 4 that is surrounded by the known Rif R mutations. Electron density also indicated shifts and/or ordering of several beta residues interacting directly with rifampicin, including Gln390, Leu391, Gln393, Asp396, His406, Arg409 and Leu413. Only very small shifts in localized regions of the **protein** backbone were indicated. (498 pages)

L29 ANSWER 5 OF 44 BIOTECHDS
ACCESSION NUMBER: 2002-18835 BIOTECHDS
TITLE: Crystal of LuxS protein which is involved in production of autoinducer-2 for identifying modulators useful for treating e.g. infection disease, stomach cancer, stomach ulcer and other intestinal complications;
vector-mediated gene transfer, expression in host cell and

computer bioinformatic software for recombinant protein production and drug screening

AUTHOR: LEWIS H A
PATENT ASSIGNEE: STRUCTURAL GENOMIX INC
PATENT INFO: WO 2002038595 16 May 2002
APPLICATION INFO: WO 2000-US30684 3 Oct 2000
PRIORITY INFO: US 2000-729838 4 Dec 2000
DOCUMENT TYPE: Patent
LANGUAGE: English
OTHER SOURCE: WPI: 2002-519293 [55]

AB DERWENT ABSTRACT:

NOVELTY - A crystal (I) comprising LuxS **protein** (which is involved in the production of autoinducer-2 (AI-2), an intercellular signaling molecule employed in the quorum sensing pathway of various bacteria) or a functional LuxS **protein** subunit in crystalline form, is new.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the following: (1) a crystal (II) comprising a homolog of LuxS **protein** having a root mean square deviation of the alpha-carbon atoms of less than 2.0 Angstrom; (2) making (M1) (II) by mixing a volume of a solution comprising the LuxS **protein** with a volume of a reservoir solution comprising a precipitant and incubating the mixture obtained over the reservoir solution in a closed container, under conditions suitable for crystallization until the crystal forms; (3) determining (M2) the three-dimensional **structure** of a LuxS **protein** crystal, by providing (I) or (II) and analyzing the crystal by X-ray diffraction; (4) a machine-readable medium embedded with: (a) information that corresponds to a three-dimensional structural representation of (I) or (II); (b) molecular **structure** coordinates as shown in the specification or at least 50% of the coordinates; or (c) molecular **structure** coordinates of a **protein** molecule comprising LuxS **protein** binding pocket comprising at least three amino acids from Glu60, Arg68, Ile81 and Asp80, Ala64, His61, Tyr91, Ser9, Phe10 and Leu7, His14, Arg23, Asp40, Arg42, Met84, Cys86 and Thr88 having the **structure** coordinate as shown in the specification or by the **structure** coordinates of a binding pocket homolog where the root mean square deviation of the backbone atoms of the amino acid residues of the binding pocket and the binding pocket homolog is less than 2.0 Angstrom;(5) producing (M3) a mutant of LuxS **protein** having altered property related to LuxS **protein** by constructing a three-dimensional **structure** of LuxS **protein** having **structure** coordinates of (I)/(II); using modeling methods to identify in the three-dimensional **structure** at least one structural portion of the LuxS **protein** molecule, where an alteration in the structural portion is predicted to result in the altered property; providing a nucleic acid molecule having a modified sequence that encodes a deletion, insertion, or substitution of one or more amino acids at a position corresponding to the structural portion; and expressing the nucleic acid molecule to produce the mutant; (6) identifying (M4) a candidate binding compound capable of binding to the active site (or accessory binding site) of LuxS **protein**, by introducing into a computer program information derived from structural coordinates defining an active site (or accessory binding site) conformation of a LuxS **protein** molecule based upon three-dimensional **structure** determination comprising an active site (or accessory binding site) formed by at least the interaction of amino acids Glu, Arg, Ile and Asp where the program utilizes or displays their three-dimensional **structure**; generating a three-dimensional representation of the active site (or accessory binding site) cavity of the LuxS **protein** in the computer program; superimposing a model of the binding test compound on the model of the active site (or accessory binding site) of the LuxS **protein**; and assessing whether the test compound model fits spatially into the active site (or accessory binding site) of the LuxS

protein; (7) selecting (M5) at least one compound that potentially binds to LuxS **protein**, by: (a) constructing a three-dimensional **structure** of LuxS **protein** and selecting at least one compound which potentially binds LuxS **protein**; (b) constructing a three-dimensional **structure** of a **protein** molecule comprising a LuxS **protein** binding pocket and computationally screening several compounds using the **structure** constructed; and (c) computationally screening a three-dimensional structural representation of a molecule comprising a LuxS **protein** binding pocket an identifying those that bind; (8) designing (M6) a compound that modulates LuxS **protein** activity by providing a computer modeling program with a set of **structure** coordinates, or a three-dimensional conformation derived from them, for a molecule that comprises a binding pocket having the structural coordinates of the binding pocket of LuxS **protein**, or a binding pocket homolog; computationally building a chemical entity represented by set of **structure** coordinates and determining whether the chemical entity is a modulator expected to bind to or interfere with the molecule; (9) a compound (C1) identified, designed or made by M4, M5 and M6; (10) a pharmaceutical composition comprising C1 or its salt and a carrier; (11) obtaining structural information about a molecule or a molecular complex of unknown **structure** by crystallizing the molecule or molecular complex; generating an x-ray diffraction pattern from the crystallized molecule or molecular complex and using a molecular replacement method to interpret the **structure** of the molecule, where the molecular replacement method uses the **structure** coordinates as given in the specification, or its subset, or the **structure** coordinates of the binding pocket; and (12) homology modeling a LuxS **protein** homolog by: (a) **aligning** the amino acid sequence of LuxS **protein** homolog with an amino acid sequence of LuxS **protein**; (b) incorporating the sequence of homolog into a model of the **structure** of LuxS **protein** ; (c) subjecting the preliminary model to **energy minimization** to yield an **energy minimized** model; and (d) remodeling regions of the **energy minimized** model where stereochemistry restraint are violated to yield a final model of the homolog.

BIOTECHNOLOGY - Preferred Crystal: (I) is preferably diffraction quality, is an apo-crystal, a native crystal, and/or is a heavy-atom derivative crystal, where LuxS is Helicobacter pylori, Haemophilus influenze or Deinococcus radiodurans LuxS, or a mutant which is selenomethionine, selenocysteine mutant, conservative mutant, truncated or extended mutant. (I) is characterized by a set of **structure** coordinate that is substantially similar to the set of **structure** coordinates as given in the specification. (II) is produced by mixing a volume of a solution comprising the LuxS **protein** with a volume of a reservoir solution comprising a precipitant and incubating the mixture obtained over the reservoir solution in a closed container, under conditions suitable for crystallization until the crystal forms. **Protein** co-ordinate data is given in the patent specification. Preferred Method: In M3, the altered activity of LuxS **protein** is preferably altered binding activity or immunogenicity, where an epitope is altered. In M4, the structural coordinates correspond to the liganded or unliganded LuxS **protein**, and the binding compound is a LuxS inhibitor. M5 further comprises screening a library of compounds. The binding pocket comprises at least three amino acids from Glu60, Arg68, Ile81 and Asp80, Ala64, His61, Tyr91, Ser9, Phe10 and Leu7, His14, Arg23, Asp40, Arg42, Met84, Cys86 and Thr88 having the **structure** coordinate as shown in the specification or a molecule comprising a binding pocket homolog where the root mean square deviation of the backbone atoms of the amino acid residues of the binding pocket and the binding pocket homolog is less than 2.0 Angstrom. The method comprises determining whether the compound potentially binds to the molecule by performing a fitting operation between the compound and a

binding pocket of the molecule or molecular complex, and computationally analyzing the results of the fitting operation to quantify the association between, or the interference with, the compound and the binding pocket.

ACTIVITY - Antibacterial; Cytostatic; Antiulcer. No supporting biological data is given.

MECHANISM OF ACTION - LuxS **protein** modulator (claimed). No supporting biological data is given.

USE - C1 is useful for modulating LuxS **protein** activity (claimed), useful for treating e.g. infection disease, stomach cancer, stomach ulcer and other intestinal complications.

ADMINISTRATION - C1 is administered through oral, buccal, sublingual, rectal, transdermal, vaginal, transmucosal, nasal or intestinal administration, parenteral delivery, including intramuscular, subcutaneous, intramedullar injections, as well as intrathecal, direct intraventricular, intravenous, intraperitoneal, intranasal or intraocular injections. Dosage of C1 is for 0.01-1000 (preferably 10-30) mg/day.

EXAMPLE - An open-reading frame for LuxS was amplified from Helicobacter pylori (Hp-ATCC43504D) genomic DNA by the polymerase chain reaction (PCR) using the following primers: Forward primer GGATTTCACATATGAAAATGAATGTAGAGAGTTTC, Reverse Primer: GTTCGGATCCAACCCCCACTTCAGACC. The PCR product (456 bp expected) was digested with NdeI and BamHI, electrophoresed on a 1% agarose gel in TBE buffer and the appropriate size band was excised from the gel and eluted using a standard gel extraction kit. The eluted DNA was ligated overnight with T4 DNA ligase at 16 degreesC into pSB3, previously digested with NdeI and BamHI. The vector pSB3 was a modified version of pET26b where the following sequence had been inserted into the BamHI siteL GGATCCCACCACCACCACCACCACTGAGATCC. The resulting sequence of the gene after being ligated into the vector, from the Shine-Dalgarno sequence through the stop site and the original BamHI, site was as follows: AAGGAGGAGATATACATATG(open reading frame (ORF))GGATCCCACCACCACCACCACCACTGA GATCC. The LuxS expressed using this vector had 8 amino acids to the C-terminal end (Gly-Ser-His-His-His-His-His-His). Plasmids containing ligated inserts were transformed into chemically competent Escherichia coli such as Top 10 cells. Colonies were then screened for inserts in the correct orientation and miniprepped. The miniprep DNA was transformed into BL21 (DE3) Active Motif cells and plated onto petri dishes containing Luria-Bertani medium (LB) agar with 30 mug/ml of kanamycin. Isolated, single colonies were grown to mid-log phase and stored at -80 degrees Centigrade in LB containing 15% glycerol. LuxS containing selenomethionine was overexpressed in Escherichia coli and the cultures were allowed to ferment overnight and the LuxS was purified. For crystals of Helicobacter pylori from which the molecular **structure** coordinates of were obtained, it had been found that a hanging drop containing 1 microlitre of LuxS polypeptide 5 mg/mL in 10 mM HEPES pH 7.5, 150 mM NaCl, 1 mM betaME, 10 mM methionine, and 1 microlitre reservoir solution 32% (w/v) PEG1000, 200 mM ammonium sulfate, 2 mM beta-mercaptoethanol, and 100 mM MES, pH 5.75 in a sealed container containing 500 microlitres reservoir solution, incubated for 3-7 days at 20 degrees Centigrade provide diffraction quality crystals.(473 pages)

L29 ANSWER 6 OF 44 BIOTECHDS COPYRIGHT 2003 THOMSON DERWENT/ISI on STN
ACCESSION NUMBER: 2003-02336 BIOTECHDS
TITLE: Molecule or molecular complex useful for obtaining structural information of a molecule of unknown **structure**, comprises a portion of Staphylococcus aureus FemA or FemA-like substrate binding surface/binding sites; database, bioinformatic software and bioinformatic hardware for protein **structure** determination
AUTHOR: BENSON T E; PRINCE D B
PATENT ASSIGNEE: BENSON T E; PRINCE D B
PATENT INFO: US 2002072105 13 Jun 2002
APPLICATION INFO: US 2001-932474 17 Aug 2001

PRIORITY INFO: US 2001-932474 17 Aug 2001; US 2000-226239 17 Aug 2000
DOCUMENT TYPE: Patent
LANGUAGE: English
OTHER SOURCE: WPI: 2002-673330 [72]

AB DERWENT ABSTRACT:

NOVELTY - A molecule or molecular complex (I) having a portion of a Staphylococcus aureus (Sa) FemA or FemA-like substrate binding surface (SBS)/binding sites (BS) has amino acids, given in specification, where SBS/BS is defined by a set of points with a root mean square deviation less than 1.5 Angstrom, from points representing backbone atoms of amino acids represented by **structure** coordinates of Sa molecules.

DETAILED DESCRIPTION - A new molecule or molecular complex (I) comprises a portion of Sa FemA or Sa FemA-like SBS or BS, where SBS or BS comprises amino acids, given in the specification, and SBS or BS is defined by a set of points having a root mean square deviation of less than about 1.5Angstrom from points representing the backbone atoms of the amino acids represented by the **structure** coordinates (SC) for molecules of Sa, given in the specification. INDEPENDENT CLAIMS are also included for the following: (1) a molecule or molecular complex (II) that is structurally homologous to Sa FemA molecule or molecular complex, where the Sa FemA molecule or molecular complex, is represented by a portion of SC; (2) a scalable three-dimensional configuration (III) of points, where a portion of the points is derived from SC of a portion of Sa FemA molecule or molecular complex, or a portion of a molecule or molecular complex that is structurally homologous to Sa FemA molecule or molecular complex and comprises a FemA or FemA-like binding site or SBS; (3) a machine-readable data storage medium comprising a data storage material encoded with a machine readable data which, when using a machine programmed with instructions for using the data, displays a graphical three-dimensional representing (I) or (II), or comprises a data storage material encoded with a set of machine readable data which, when combined with a second set of machine readable data, using a machine programmed with instructions for using the first and second set of data, determines a portion of the **structure** coordinates corresponding to the second set of machine readable data, where the first set of data comprises a Fourier transform of a portion of SC for Sa FemA, and the second set of data comprises an X-ray diffraction pattern of a molecule or molecular complex of unknown **structure**; (4) homology modeling of a Sa FemA homolog, comprises: (a) **aligning** the amino acid sequence of Sa FemA homolog with an amino acid sequence of Sa FemA (comprising a sequence of 414 amino acids, given in the specification) and incorporating the sequence of a Sa FemA homolog into a model of Sa FemA formed from SC to yield a preliminary model of Sa FemA homolog; (b) subjecting the preliminary model to **energy minimization** to yield a **energy minimized** model; and (c) remodeling regions of the **energy minimized** model, where stereochemistry restrains are violated to yield a final model of Sa FemA homolog; (5) a computer-assisted method (M1) for identifying a potential modifier of Sa FemA activity, comprising: (a) supplying a computer modeling application with a set of **structure** coordinates of a molecule or a molecular complex, where the molecule or molecular complex comprises a portion of Sa FemA or Sa FemA-like SBS or BS, and the SBS or binding site comprises the amino acids, given in the specification; (b) supplying the computer modeling application with a set of **structure** coordinates of a chemical entity; (c) optionally, evaluating the potential binding or interfering interactions between the chemical entity and SBS or binding site of the molecule or molecular complex, and structurally modifying the chemical entity to yield a set of **structure** coordinates for a modified chemical entity, or computationally building a chemical entity represented by a set of **structure** coordinates; and (d) determining whether the chemical entity is expected to bind to or interfere with the molecule or molecular complex, where binding to or interfering with the molecule or molecular complex is indicative of

potential modification of Sa FemA activity; (6) making (M2) a potential modifier of Sa FemA activity, comprising chemically or enzymatically synthesizing a chemical entity to yield a potential modifier of Sa FemA activity, where the chemical entity has been identified by M1; (7) a potential modifier (IV) of Sa FemA activity identified, designed or made by M1 or M2; (8) a composition (C) comprising (IV); (9) a pharmaceutical composition (PC) comprising (IV) or its salt and a carrier; (10) crystallizing a Sa FemA molecule or molecular complex, by preparing a purified Sa FemA at a concentration of 1 - 50 mg/ml and crystallizing Sa FemA from a solution comprising 1 - 50 weight % (wt.%) of polyethylene glycol (PEG), 0 - 50 wt.% glycerol, 0 - 1 M NaCl, 0 - 40 wt.% of dimethyl sulfoxide (DMSO), 100 mM - 1 M Ca(OAc)2, and/or MgCl2, and buffered to a pH of 7 to 10; and (11) a crystal (V) of Sa FemA.

WIDER DISCLOSURE - Also disclosed are: (1) computational screening of small molecule databases for chemical entities or compounds that bind in whole, or in part, to Sa FemA or Sa FemA-like SBS or BS; and (2) a magnetic storage media including (III).

BIOTECHNOLOGY - Preferred Crystal: (V) has the orthorhombic space group symmetry P212121 and comprises a unit cell having dimensions, a, b and c, where a is about 40 - 70 Angstrom, b is 75 - 105 Angstrom, and c is 95 - 125 Angstrom, and alpha=beta=gamma=90 degrees, or comprises atoms arranged in a spatial relationship represented by SC. (V) has amino acids having the sequence of (S1), with the proviso that a methionine is replaced with selenomethionine.

ACTIVITY - None given.

MECHANISM OF ACTION - Sa FemA activity inhibitor (claimed). No biological data is given.

USE - (I) is useful for obtaining structural information about a molecule or a molecular complex of unknown **structure**, by: (a) crystallizing (I); (b) generating an X-ray diffraction pattern from the crystallized molecule or molecular complex; and (c) applying a portion of SC to the X-ray diffraction pattern to generate a three-dimensional electron density map of a portion of the molecular or molecular complex whose **structure** is unknown (claimed). A potential modifier (IV) of Sa FemA activity is useful for preventing and/or treating Sa FemA mediated diseases i.e., used in chronic or acute therapy. A crystal (V) of Sa FemA is useful for solving the **structure** of other molecules or molecular complexes and for identifying and/or designing modifiers of FemA activity, and for rational drug designing by probing Sa FemA crystals with molecules including a variety of different functional groups to determine sites for interaction between candidate Sa FemA modifiers and the **protein**. (V) is useful in X-ray crystallographic analysis.

ADMINISTRATION - Administered at a dose of 0.01 - 100 (preferably 0.5 - 75) mg/kg body weight, through oral, parenteral (subcutaneous, intracutaneous, intravenous, intramuscular, intrarticular, intrasynovial, intrasternal, intrathecal, intralesional or intracranial), inhalation, topical, rectal, nasal, buccal or vaginal routes.

EXAMPLE - Methionine incorporated Staphylococcus aureus Fem A was obtained in 50 mM ethanolamine, 1 mM dithiotheritol (DTT), pH 10.0. The **protein** was concentrated to 12 mg/ml. The concentrated sample was used to begin screening FemA in the crystallization screening library. The crystals were rod shaped, and were 100 - 250 micrometers long, and 30 - 50 micrometers thick. Optimization around these conditions was started with the Hampton follow-up library and the 2 crystals repeated, growing to 200 micrometers x 20 micrometers. These crystals were taken to the synchrotron for data collection, where they diffracted to 2.7 Angstrom. Another crystal form, which produced thicker rods, was found during a selenomethionine incorporated S. aureus Fem A screen. This condition, Wizard screen I condition 46 (10 % polyethylene glycol (PEG) 8000, 0.2 M Ca(OAc)2, 0.1 M imidazole, pH 8.0), produced crystals suitable for diffraction studies from the screen. The crystal was soaked in a suitable cryoprotectant agent and stored. One selenomethionine multiple anomalous dispersion (MAD) experiment was performed (2.1 Angstrom resolution) using

three different wavelengths (2.10 Angstrom, 2.06 Angstrom and 2.09 Angstrom). Each of these individual data sets was indexed and integrated separately. The data sets for each experiment was scaled to each other using the program SCALEIT (RTM) in the CCP4 Program Suite (RTM). Patterson maps revealed six selenium sites whose locations were determined by direct methods using SHELX (RTM). Two pairs of three sites each were tested for authenticity by their ability to generate phases which could identify the other pair of sites in anomalous difference Fourier calculations. A subsequent site was identified by anomalous difference Fourier methods. The seven sites accounted for all of the methionines in the **protein** including the N-terminal methionine.
All heavy atom parameter refinement and phasing calculations were carried out with MLPHARE (RTM) by treating the remote wavelength as native and the edge and peak wavelengths as derivatives. The phases were subsequently subjected to solvent flattening using the program DM. The multiple anomalous dispersion phased electron density map was exceptionally clear. The initial placement of the C(alpha) backbone and correlation between the sequence and the main chain was done using the X-AutoFit module in Quanta (RTM). Because of the high quality of the phases, water molecules were added based on the MAD phased electron density map. Before refinement, the starting R-factor/Free R-factor was 39.5 %/40.5 %. One cycle of positional refinement, torsion angle dynamics refinement, and individual B factor refinement with a bulk solvent correction led to significant improvement in the model (R-factor/Free R-factor = 24 %/28.2 %). The rapid drop in the R-factor during the first cycle of refinement reflected the high quality phases that were determined and used to calculate the initial electron density map. Three more cycles of refinement and rebuilding led to the model R-factor/Free R-factor = 20.5 %/24.5 %). All refinement cycles were carried out with XPLOR98 (RTM), incorporating bulk solvent correction during the refinement. Stereochemistry of the model was checked using PROCHECK (RTM) revealing only two residues in disallowed regions of Ramachandran plot. (37 pages)

L29 ANSWER 7 OF 44 CAPLUS COPYRIGHT 2003 ACS on STN

ACCESSION NUMBER: 2002:121268 CAPLUS
DOCUMENT NUMBER: 137:75473
TITLE: **Protein structure** prediction using a combination of sequence-based **alignment**, constrained **energy minimization**, and structural **alignment**
AUTHOR(S): Standley, Daron M.; Eyrich, Volker A.; An, Yuling; Pincus, David L.; Gunn, John R.; Friesner, Richard A.
CORPORATE SOURCE: Schrodinger Inc., Jersey City, NJ, USA
SOURCE: Proteins: Structure, Function, and Genetics (2002), Volume Date 2001, (Suppl. 5), 133-139
CODEN: PSFGEY; ISSN: 0887-3585
PUBLISHER: Wiley-Liss, Inc.
DOCUMENT TYPE: Journal
LANGUAGE: English
AB We present a novel approach to protein **structure** prediction in which fold recognition techniques are combined with ab initio folding methods. Based on the predicted secondary **structure**, one of two different protocols is followed. For mostly .alpha.-proteins, global optimization and sampling of a statistical energy function is used to generate many low-energy structures; these structures are then screened against a fold library. Any structural matches are then selected for further refinement. For proteins predicted to have significant .beta.-content, sequence and secondary **structure**-based **alignment** is used to identify candidate templates; spatial constraints are then extd. from these templates and used, along with the statistical energy function, in the global sampling and optimization program. Successes and failures of both protocols are discussed.
REFERENCE COUNT: 16 THERE ARE 16 CITED REFERENCES AVAILABLE FOR THIS

RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L29 ANSWER 8 OF 44 CAPLUS COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER: 2001:893412 CAPLUS
DOCUMENT NUMBER: 136:130639
TITLE: Solution **structure** of a Kunitz-type chymotrypsin inhibitor isolated from the elapid snake Bungarus fasciatus
AUTHOR(S): Chen, Chinpan; Hsu, Chun-Hua; Su, Ning-Yuan; Lin, Yu-Ching; Chiou, Shyh-Horng; Wu, Shih-Hsiung
CORPORATE SOURCE: Institutes of Biomedical Sciences, Academia Sinica, Taipei, 115, Taiwan
SOURCE: Journal of Biological Chemistry (2001), 276(48), 45079-45087
CODEN: JBCHA3; ISSN: 0021-9258
PUBLISHER: American Society for Biochemistry and Molecular Biology
DOCUMENT TYPE: Journal
LANGUAGE: English

AB Bungarus fasciatus fraction IX (BF9), a chymotrypsin inhibitor, consists of 65 amino acid residues with three disulfide bridges. It was isolated from the snake venom of B. fasciatus by ion-exchange chromatog. and belongs to the bovine pancreatic trypsin inhibitor (BPTI)-like superfamily. It showed a dissocn. const. of 5.8 .times. 10-8 M with .alpha.-chymotrypsin as measured by a BIAcore binding assay system. The isothermal titrn. calorimetry revealed a 1:1 binding stoichiometry between this inhibitor and chymotrypsin and apparently no binding with trypsin. We further used CD and NMR to det. the soln. **structure** of this venom-derived chymotrypsin inhibitor. The three-dimensional NMR soln. structures of BF9 were detd. on the basis of 582 restraints by simulated annealing and **energy minimization** calcns. The final set of 10 NMR structures was well defined, with av. root mean square deviations of 0.47 .ANG. for the backbone atoms in the secondary **structure** regions and 0.86 .ANG. for residues The side chains of Phe23, Tyr24, Tyr25, Phe35, and Phe47 exhibited many long-range nuclear Overhauser effects and were the principal components of the hydrophobic core in BF9. To gain insight into the **structure**-function relationships among proteins in the BPTI-like super-family, we compared the three-dimensional **structure** of BF9 with three BPTI-like proteins that possess distinct biol. functions. These proteins possessed similar secondary **structure** elements, but the loop regions and .beta.-turn were different from one another. Based on residues at the functional site of each protein, we suggest that the flexibility, rigidity, and variations of the amino acid residues in both the loop and .beta.-turn regions are related to their biol. functions.

REFERENCE COUNT: 38 THERE ARE 38 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L29 ANSWER 9 OF 44 MEDLINE on STN DUPLICATE 1
ACCESSION NUMBER: 2001468795 MEDLINE
DOCUMENT NUMBER: 21404147 PubMed ID: 11513239
TITLE: Three-dimensional quantitative **structure**-activity relationship for several bioactive peptides searched by a convex hull-comparative molecular field analysis approach.
AUTHOR: Lin T H; Lin J J
CORPORATE SOURCE: Department of Life Science, National Tsing Hua University, Hsinchu, Taiwan, ROC.. lslth@life.nthu.edu.tw
SOURCE: COMPUTERS AND CHEMISTRY, (2001 Sep) 25 (5) 489-98.
Journal code: 7607706. ISSN: 0097-8485.
PUB. COUNTRY: England: United Kingdom
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 200110

ENTRY DATE: Entered STN: 20010830
Last Updated on STN: 20011008
Entered Medline: 20011004

AB Three-dimensional (3D) convex hulls are computed for theoretically generated structures of a group of 18 bioactive tachykinin peptides. The number of peptides treated as a training set is 14, whereas that treated as a test set is four. The frequency of atoms of the same atomic type lying at the vertices of all the hulls computed for all the structures in a structural set is counted. Vertex atoms with non-zero frequency counted are collected together as a set of commonly exposed groups. These commonly exposed atoms are then treated as a set of correspondences for **aligning** all the other 13 structures in a structural set against a common template, which is the **structure** of the most potent **peptide** in the set using the FIT module of the SYBYL 6.6 program. Each aligned structural set is then analyzed by the comparative molecular field analysis (CoMFA) module using the C.3 probe having a charge of +1.0. The corresponding cross-validated r2 values range from -0.99 to 0.57 for a number of 73 structural sets studied. The comparative molecular similarity indices analysis (CoMSIA) module within the SYBYL 6.6 package is also used to analyze some of these aligned structural sets. Although the CoMSIA results are in accord with those of CoMFA, it is also found that the CoMFA results of several structural sets can be improved somewhat for conformations of the structures in the sets that are adjusted by constraint **energy minimization** and then constraint molecular dynamics simulation runs using distance constraints derived from some commonly exposed groups determined for them. This result further implies that the convex hull-CoMFA is a feasible approach to screen the bioactive conformations for molecules of this class.

L29 ANSWER 10 OF 44 BIOSIS
DUPLICATE 2

ACCESSION NUMBER: 2002:95420 BIOSIS
DOCUMENT NUMBER: PREV200200095420
TITLE: Prediction of potential toxicity and side effect protein targets of a small molecule by a ligand-protein inverse docking approach.
AUTHOR(S): Chen, Y. Z. (1); Ung, C. Y.
CORPORATE SOURCE: (1) Department of Computational Science, National University of Singapore, Blk S17, 3 Science Drive 2, Level 7, Singapore, 117543: yzchen@cz3.nus.edu.sg Singapore
SOURCE: Journal of Molecular Graphics & Modelling, (2001) Vol. 20, No. 3, pp. 199-218. http://www.elsevier.nl/inca/publications/store/5/2/5/0/1/2/index.htt. print.
ISSN: 1093-3263.
DOCUMENT TYPE: Article
LANGUAGE: English

AB Determination of potential drug toxicity and side effect in early stages of drug development is important in reducing the cost and time of drug discovery. In this work, we explore a computer method for predicting potential toxicity and side effect **protein** targets of a small molecule. A ligand-**protein** inverse docking approach is used for computer-automated search of a **protein** cavity database to identify **protein** targets. This database is developed from **protein** 3D structures in the **protein** data bank (PDB). Docking is conducted by a procedure involving multiple conformer shape-matching **alignment** of a molecule to a cavity followed by molecular-mechanics torsion optimization and **energy minimization** on both the molecule and the **protein** residues at the binding region. Potential **protein** targets are selected by evaluation of molecular mechanics energy and, while applicable, further analysis of its binding competitiveness against other ligands that bind to the same receptor site in at least one PDB entry. Our results on several drugs show that 83% of the experimentally known toxicity and side effect targets for these drugs are predicted. The

computer search successfully predicted 38 and missed five experimentally confirmed or implicated **protein** targets with available **structure** and in which binding involves no covalent bond. There are additional 30 predicted targets yet to be validated experimentally. Application of this computer approach can potentially facilitate the prediction of toxicity and side effect of a drug or drug lead.

L29 ANSWER 11 OF 44 MEDLINE on STN DUPLICATE 3
ACCESSION NUMBER: 2002113831 MEDLINE
DOCUMENT NUMBER: 21824353 PubMed ID: 11835490
TITLE: **Protein structure** prediction using a combination of sequence-based **alignment**, constrained **energy minimization**, and structural **alignment**.
AUTHOR: Standley D M; Eyrich V A; An Y; Pincus D L; Gunn J R; Friesner R A
CORPORATE SOURCE: Schrodinger Inc., Jersey City, New Jersey, USA.
CONTRACT NUMBER: GM52018 (NIGMS)
RR06892 (NCRR)
SOURCE: PROTEINS, (2001) Suppl 5 133-9.
Journal code: 8700181. ISSN: 0887-3585.
PUB. COUNTRY: United States
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 200208
ENTRY DATE: Entered STN: 20020216
Last Updated on STN: 20020829
Entered Medline: 20020827

AB We present a novel approach to protein **structure** prediction in which fold recognition techniques are combined with ab initio folding methods. Based on the predicted secondary **structure**, one of two different protocols is followed. For mostly alpha proteins, global optimization and sampling of a statistical energy function is used to generate many low-energy structures; these structures are then screened against a fold library. Any structural matches are then selected for further refinement. For proteins predicted to have significant beta-content, sequence and secondary **structure**-based **alignment** is used to identify candidate templates; spatial constraints are then extracted from these templates and used, along with the statistical energy function, in the global sampling and optimization program. Successes and failures of both protocols are discussed. Copyright 2002 Wiley Liss, Inc.

L29 ANSWER 12 OF 44 EMBASE COPYRIGHT 2003 ELSEVIER SCI. B.V. on STN
ACCESSION NUMBER: 2002049163 EMBASE
TITLE: **Protein structure** prediction using a combination of sequence-based **alignment**, constrained **energy minimization**, and structural **alignment**.
AUTHOR: Standley D.M.; Eyrich V.A.; An Y.; Pincus D.L.; Gunn J.R.; Friesner R.A.
CORPORATE SOURCE: R.A. Friesner, Department of Chemistry, Center for Biomolecular Simulation, Columbia University, New York, NY 10027, United States. rich@chem.columbia.edu
SOURCE: Proteins: Structure, Function and Genetics, (2001) 45/SUPPL. 5 (133-139).
Refs: 16
ISSN: 0887-3585 CODEN: PSFGEY
COUNTRY: United States
DOCUMENT TYPE: Journal; Article
FILE SEGMENT: 029 Clinical Biochemistry
LANGUAGE: English
SUMMARY LANGUAGE: English

AB We present a novel approach to **protein structure** prediction in which fold recognition techniques are combined with ab initio folding methods. Based on the predicted secondary **structure**, one of two different protocols is followed. For mostly .alpha.-proteins, global optimization and sampling of a statistical energy function is used to generate many low-energy structures; these structures are then screened against a fold library. Any structural matches are then selected for further refinement. For proteins predicted to have significant .beta.-content, sequence and secondary **structure**-based **alignment** is used to identify candidate templates; spatial constraints are then extracted from these templates and used, along with the statistical energy function, in the global sampling and optimization program. Successes and failures of both protocols are discussed. .COPYRGT. 2002 Wiley-Liss, Inc.

L29 ANSWER 13 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN DUPLICATE 4

ACCESSION NUMBER: 2000:387273 BIOSIS
DOCUMENT NUMBER: PREV200000387273
TITLE: The highly refined solution **structure** of the cytotoxic ribonuclease alpha-sarcin reveals the structural requirements for substrate recognition and ribonucleolytic activity.
AUTHOR(S): Perez-Canadillas, Jose Manuel; Santoro, Jorge; Campos-Olivas, Ramon; Lacadena, Javier; Martinez del Pozo, Alvaro; Gavilanes, Jose G.; Rico, Manuel; Bruix, Marta (1)
CORPORATE SOURCE: (1) Consejo Superior de Investigaciones Cientificas, Instituto de Estructura de la Materia, Serrano 119, 28006, Madrid Spain
SOURCE: Journal of Molecular Biology, (16 June, 2000) Vol. 299, No. 4, pp. 1061-1073. print.
ISSN: 0022-2836.
DOCUMENT TYPE: Article
LANGUAGE: English
SUMMARY LANGUAGE: English
AB alpha-Sarcin selectively cleaves a single phosphodiester bond in a universally conserved sequence of the major rRNA, that inactivates the ribosome. The elucidation of the three-dimensional solution **structure** of this 150 residue enzyme is a crucial step towards understanding alpha-sarcin's conformational stability, ribonucleolytic activity, and its exceptionally high level of specificity. Here, the solution **structure** has been determined on the basis of 2658 conformationally relevant distances restraints (including stereoespecific assignments) and 119 torsional angular restraints, by nuclear magnetic resonance spectroscopy methods. A total of 60 converged structures have been computed using the program DYANA. The 47 best DYANA structures, following restrained **energy minimization** by GROMOS, represent the solution **structure** of alpha-sarcin. The resulting average pairwise root-mean-square-deviation is 0.86 ANG for backbone atoms and 1.47 ANG for all heavy atoms. When the more variable regions are excluded from the analysis, the pairwise root-mean-square deviation drops to 0.50 ANG and 1.00 ANG, for backbone and heavy atoms, respectively. The alpha-sarcin **structure** is similar to that reported for restrictocin, although some differences are clearly evident, especially in the loop regions. The average rmsd between the structurally **aligned** backbones of the 47 final alpha-sarcin structures and the crystal **structure** of restrictocin is 1.46 ANG. On the basis of a docking model constructed with alpha-sarcin solution **structure** and the crystal **structure** of a 29-nt RNA containing the sarcin/ricin domain, the regions in the **protein** that could interact specifically with the substrate have been identified. The structural elements that account for the specificity of RNA recognition are located in two separate regions of the **protein**. One is composed by residues 51 to 55 and loop 5, and the other region, located

more than 11 ANG away in the **structure**, is the positively charged segment formed by residues 110 to 114.

L29 ANSWER 14 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN DUPLICATE 5

ACCESSION NUMBER: 2000:376238 BIOSIS
DOCUMENT NUMBER: PREV200000376238
TITLE: Homology-based molecular modelling of PLP-dependent histidine decarboxylase from Morganella morganii.
AUTHOR(S): Tahanejad, Fatemeh Sadat; Naderi-Manesh, Hossein (1); Habibinejad, Bahram; Mahmoudian, Massoud
CORPORATE SOURCE: (1) School of Sciences, Tarbiat Modarres University, Tehran Iran
SOURCE: European Journal of Medicinal Chemistry, (June, 2000) Vol. 35, No. 6, pp. 567-576. print.
ISSN: 0223-5234.
DOCUMENT TYPE: Article
LANGUAGE: English
SUMMARY LANGUAGE: English

AB The 3-D structural information is a prerequisite for a rational ligand design. In the absence of experimental data, model building on the basis of a known 3-D **structure** of a homologous **protein** is at present the only reliable method to obtain structural information. A homology model building study of the pyridoxal 5'-phosphate (PLP)-dependent histidine decarboxylase from Morganella morganii (HDC-MM) has been carried out based on the crystal **structure** of the aspartate aminotransferase from Escherichia coli (AAT-EC). The primary sequences of AAT-EC and HDC-MM were **aligned** by automated **alignment** procedure. A 3-D model of HDC-MM was constructed by copying the coordinates of the residues from the crystal **structure** of AAT-EC into the corresponding residues in HDC-MM. After **energy -minimization** of the resulting 3-D model of HDC-MM, possible active site residues were identified by fitting the substrate (l-histidine) into the proposed active-site. In our model, several residues, which have an important role in the AAT-EC active-site, are located in positions spatially identical to those in AAT-EC **structure**. The back-bone of the modelled active site pocket is constructed by residues; Gly-92, Gly-93, Thr-93, Ser-115, Asp-200, Ala-202, Ser-229 and Lys-232 together with residues Asn-8, His-119, Thr-171, His-198, Leu-203, His-231, Ser-236 and Ile-238. In the ligand binding site, it appears that the HDC-MM model will position l-histidine (substrate) in the area consisting of the residues; Glu-29, Ser-30, Leu-38, His-231 and Lys-232. The nitrogen atom of the imidazole ring (N2) of the substrate is predicted to interact with the carboxylate group of Ser-30. The alpha-carboxylate of histidine points toward the Lys-232 to have electrostatic interaction with its side chain nitrogen atom (NZ). In conclusion, this combination of sequence and 3-D structural homology between AAT-EC and HDC-MM model could provide insight in assigning the probable active site residues.

L29 ANSWER 15 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN DUPLICATE 6

ACCESSION NUMBER: 2000:397641 BIOSIS
DOCUMENT NUMBER: PREV200000397641
TITLE: Comparative molecular modeling on 3D-**structure** of opioid receptor-like 1 receptor.
AUTHOR(S): Huang Xiao-Qin; Jiang Hua-Liang (1); Luo Xiao-Min; Chen Kai-Xian (1); Zhu You-Cheng; Ji Ru-Yun; Cao Yang
CORPORATE SOURCE: (1) Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai, 200031 China
SOURCE: Acta Pharmacologica Sinica, (June, 2000) Vol. 21, No. 6, pp. 529-535. print.
ISSN: 0253-9756.
DOCUMENT TYPE: Article

LANGUAGE: English
SUMMARY LANGUAGE: Chinese; English

AB AIM: To build the three-dimensional **structure** of opioid receptor-like 1 (ORL1) receptor. METHODS: Structural elements of ORL1 receptor were predicted from sequence **alignments** of opioid and related receptors of G **protein**-coupled receptor (GPCR) based on (i) the consensus, biophysical interpretations of **alignment** -derived properties, and (ii) tertiary structural homology to frog rhodopsin; The extracellular loops of ORL1 were built by self-constructed database searching based on geometrical constraints; initial model was refined computationally with **energy minimization** by molecular mechanics method. RESULTS: The calculated **structure** of ORL1 receptor has clusters of hydrogen bonds existing in inter-helices and extracellular loops; the ORL1 receptor has a possible ligand-binding "crevice" situated on the extraside of the transmembrane domains between helices 3, 5, 6, and 7, which is partially covered by the extracellular loop 2 (EL-2); The binding cavity may consist of a "highly conserved region" involving the residues of Asp130, Tyr131, and an outer "conservatively variable region" containing the residues near the interface of transmembrane (TM) helices-EL loops; The molecular model obtained is qualitatively consistent with ligand affinities, hybrid **peptide** studies, and other experimental data. CONCLUSION: The structural model of ORL1 receptor from this study is helpful for clarifying experimental observations of ligands interacting with opioid receptors, and for designing new biological investigations.

L29 ANSWER 16 OF 44 MEDLINE on STN DUPLICATE 7

ACCESSION NUMBER:	2000146255 MEDLINE
DOCUMENT NUMBER:	20146255 PubMed ID: 10679615
TITLE:	The third-dimensional **structure** of the complex between an Fv antibody fragment and an analogue of the main immunogenic region of the acetylcholine receptor: a combined two-dimensional NMR, homology, and molecular modeling approach.
AUTHOR:	Kleinjung J; Petit M C; Orlewski P; Mamalaki A; Tzartos S J; Tsikaris V; Sakarellos-Daitsiotis M; Sakarellos C; Marraud M; Cung M T
CORPORATE SOURCE:	Laboratoire de Chimie-Physique Macromoleculaire, UMR 7568 CNRS-INPL, ENSIC, BP 451, 54001 Nancy Cedex, France.
SOURCE:	BIOPOLYMERS, (2000 Feb) 53 (2) 113-28. Journal code: 0372525. ISSN: 0006-3525.
PUB. COUNTRY:	United States
DOCUMENT TYPE:	Journal; Article; (JOURNAL ARTICLE)
LANGUAGE:	English
FILE SEGMENT:	Priority Journals
ENTRY MONTH:	200003
ENTRY DATE:	Entered STN: 20000330 Last Updated on STN: 20000330 Entered Medline: 20000321

AB Binding of autoantibodies to the acetylcholine receptor (AChR) plays a major role in the autoimmune disease Myasthenia gravis (MG). In this paper, we propose a **structure** model of a putative immunocomplex that gives rise to the reduction of functional AChR molecules during the course of MG. The model complex consists of the [G(70), Nle(76)] decapeptide analogue of the main immunogenic region (MIR), representing the major antigenic epitope of AChR, and the single chain Fv fragment of monoclonal antibody 198, a potent MG autoantibody. The **structure** of the complexed decapeptide antigen [G(70), Nle(76)]MIR was determined using two-dimensional nmr, whereas the antibody **structure** was derived by means of homology modeling. The final complex was constructed using calculational docking and molecular dynamics. We termed this approach "directed modeling," since the known peptide **structure** directs the prestructured antibody binding site to its final conformation. The independently derived structures of the peptide antigen and antibody

binding site already showed a high degree of surface complementarity after the initial docking calculation, during which the peptide was conformationally restrained. The docking routine was a soft algorithm, applying a combination of Monte Carlo simulation and **energy minimization**. The observed shape complementarity in the docking process suggested that the **structure** assessments already led to anti-idiotypic conformations of peptide antigen and antibody fragment. Refinement of the complex by dynamic simulation yielded improved surface adaptation by small rearrangements within antibody and antigen. The complex presented herein was analyzed in terms of antibody-antigen interactions, properties of contacting surfaces, and segmental mobility. The structural requirements for AChR complexation by autoantibodies were explored and compared with experimental data from alanine scans of the MIR peptides. The analysis revealed that the N-terminal loop of the **peptide structure**, which is indispensable for antibody recognition, **aligns** three hydrophobic groups in a favorable arrangement leading to the burial of 40% of the **peptide** surface in the binding cleft upon complexation. These data should be valuable in the rational design of an Fv mutant with much improved affinity for the MIR and AChR to be used in therapeutic approaches in MG.

L29 ANSWER 17 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
DUPLICATE 8

ACCESSION NUMBER: 2000:70501 BIOSIS
DOCUMENT NUMBER: PREV200000070501
TITLE: 3D modeling, ligand binding and activation studies of the cloned mouse delta, mu and kappa opioid receptors.
AUTHOR(S): Filizola, Marta; Laakkonen, Liisa; Loew, Gilda H. (1)
CORPORATE SOURCE: (1) Molecular Research Institute, 2495 Old Middlefield Way, Mountain View, CA USA
SOURCE: Protein Engineering, (Nov., 1999) Vol. 12, No. 11, pp. 927-942.
ISSN: 0269-2139.
DOCUMENT TYPE: Article
LANGUAGE: English
SUMMARY LANGUAGE: English
AB Refined 3D models of the transmembrane domains of the cloned delta, mu and kappa opioid receptors belonging to the superfamily of G-**protein** coupled receptors (GPCRs) were constructed from a multiple sequence **alignment** using the alpha carbon template of rhodopsin recently reported. Other key steps in the procedure were relaxation of the 3D helix bundle by unconstrained energy optimization and assessment of the stability of the **structure** by performing unconstrained molecular dynamics simulations of the energy optimized **structure**. The results were stable ligand-free models of the TM domains of the three opioid receptors. The ligand-free delta receptor was then used to develop a systematic and reliable procedure to identify and assess putative binding sites that would be suitable for similar investigation of the other two receptors and GPCRs in general. To this end, a non-selective, 'universal' antagonist, naltrexone, and agonist, etorphine, were used as probes. These ligands were first docked in all sites of the model delta opioid receptor which were sterically accessible and to which the protonated amine of the ligands could be anchored to a complementary proton-accepting residue. Using these criteria, nine ligand-receptor complexes with different binding pockets were identified and refined by **energy minimization**. The properties of all these possible ligand-substrate complexes were then examined for consistency with known experimental results of mutations in both opioid and other GPCRs. Using this procedure, the lowest energy agonist-receptor and antagonist-receptor complexes consistent with these experimental results were identified. These complexes were then used to probe the mechanism of receptor activation by identifying differences in receptor conformation between the agonist and the antagonist complex during unconstrained

dynamics simulation. The results lent support to a possible activation mechanism of the mouse delta opioid receptor similar to that recently proposed for several other GPCRs. They also allowed the selection of candidate sites for future mutagenesis experiments.

L29 ANSWER 18 OF 44 BIOSIS
DUPLICATE 9

ACCESSION NUMBER: 1999:273890 BIOSIS
DOCUMENT NUMBER: PREV199900273890
TITLE: Characterization of a type of beta-bend ribbon spiral generated by the repeating (Xaa-Yaa-Aib-Pro) motif: The solution **structure** of harzianin HC IX, a 14-residue peptaibol forming voltage-dependent ion channels.
AUTHOR(S): Segalas, Isabelle; Prigent, Yann; Davoust, Daniel; Bodo, Bernard; Rebuffat, Sylvie (1)
CORPORATE SOURCE: (1) Laboratoire de Chimie des Substances Naturelles, Museum National d'Histoire Naturelle, ESA 8041 CNRS, GDR 790 CNRS, 63 rue Buffon, 75231, Paris Cedex 05 France
SOURCE: Biopolymers, (July, 1999) Vol. 50, No. 1, pp. 71-85.
ISSN: 0006-3525.
DOCUMENT TYPE: Article
LANGUAGE: English
SUMMARY LANGUAGE: English

AB The three-dimensional solution **structure** of harzianin HC IX, a peptaibol antibiotic isolated from the fungus Tichoderma harzianum, was determined using CD, homonuclear, and heteronuclear two-dimensional nmr spectroscopy combined with molecular modeling. This 14-residue **peptide**, Ac Aib1 Asn2 Leu3 Aib4 Pro5 Ala6 Ile7 Aib8 Pro9 Iva10 Leu11 Aib12 Pro13 Leuol14 Aib, alpha-aminoisobutyric acid; Iva, isovaline; Leuol, leucinol), is a man representative of a short-sequence peptaibol class characterized by an acetylated N-terminus, a C-terminal amino alcohol, and the presence of three Aib-L-Pro motifs at positions 4-5, 8-9, and 12-13, separated by two **dipeptide** units. In spite of a lower number of residues, compared to the 18/20-residue peptaibols such as alamethicin, harzianin HC IX exhibits remarkable membrane-perturbing properties. It interacts with phospholipid bilayers, increasing their permeability and forming voltage-gated ion channels through a mechanism slightly differing from that proposed for alamethicin. Sequence-specific 1H- and 13C-nmr assignments and conformational nmr parameters (3JNHCalphaH coupling constants, quantitative nuclear Overhauser enhancement data, temperature coefficients of amide and carbonyl groups, NH-ND exchange rates) were obtained in methanol solution. Sixty structures were calculated based on 98 interproton distance restraints and 6 PHI dihedral angle restraints, using high temperature restrained molecular dynamics and **energy minimization**. Thirty-seven out of the sixty generated structures were consistent with the nmr data and were convergent. The **peptide** backbone consists in a ribbon of overlapping beta-turns twisted into a continuous spiral from Asn2 to Leuol14 and forming a 26 ANG long helix-like **structure**. This **structure** is slightly amphipathic, with the three Aib-Pro motifs **aligned** on the less hydrophobic face of the spiral where the Asn2 side chain is also present, while the more hydrophobic bulky side chains of leucines, isoleucine, isovaline, and leucinol are located on the concave side. The repetitive (Xaa-Yaa-Aib-Pro) **tetrapeptide** subunit, making up the **peptide** sequence, is characterized by four sets of (PHI,PSI) torsional angles, with the following mean values: PHIi = -90degree, PSIi = -27degree; PHIi+1 = -98degree, PSIi+1 = -17degree PHIi+2 = -49degree, PSIi+2 = -50degree; PHIi+3 = -78degree; PSIi+3 = +3degree. We term this particular **structure**, specifically occurring in the case of (Xaa-Yaa-Aib-Pro)n sequences, the (Xaa-Yaa-Aib-Pro)-beta-bend ribbon spiral. It is stabilized by 4 fwdarw 1 intramolecular hydrogen bonds and differs from both the canonical 310-helix made of a succession of type III beta-turns and from the

beta-bend ribbon spiral that has been described in the case of (Aib-Pro)n **peptide** segments.

L29 ANSWER 19 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN DUPLICATE 10

ACCESSION NUMBER: 1998:175555 BIOSIS
DOCUMENT NUMBER: PREV199800175555
TITLE: **Structure** comparison of human glioma pathogenesis-related protein GliPr and the plant pathogenesis-related protein P14a indicates a functional link between the human immune system and a plant defense system.
AUTHOR(S): Szyperski, T.; Fernandez, C.; Mumenthaler, C.; Wuthrich, K. (1)
CORPORATE SOURCE: (1) Inst. Molekularbiol. Biophysik, Eidgenossiche Tech. Hochschule-Honggerberg, CH-8093 Zurich Switzerland
SOURCE: Proceedings of the National Academy of Sciences of the United States of America, (March 3, 1998) Vol. 95, No. 5, pp. 2262-2266.
ISSN: 0027-8424.
DOCUMENT TYPE: Article
LANGUAGE: English

AB The human glioma pathogenesis-related **protein** (GliPR) is highly expressed in the brain tumor glioblastoma multiforme and exhibits 35% amino acid sequence identity with the tomato pathogenesis-related (PR) **protein** P14a, which has an important role for the plant defense system. A molecular model of GliPR was computed with the distance geometry program DIANA on the basis of a P14aGliPR sequence **alignment** and a set of 1,200 experimental NMR conformational constraints collected with P14a. The GliPR **structure** is represented by a group of 20 conformers with small residual DIANA target function values, low AMBER-energies after restrained **energy-minimization** with the program OPAL, and an average rms deviation relative to the mean of 1.6 ANG for the backbone heavy atoms. Comparison of the GliPR model with the P14a **structure** lead to the identification of a common partially solvent-exposed spatial cluster of four amino acid residues, His-69, Glu-88, Glu-110, and His-127 in the GliPR numeration. This cluster is conserved in all known plant PR proteins of class 1, indicating a common putative active site for GliPR and PR-1 proteins and thus a functional link between the human immune system and a plant defense system.

L29 ANSWER 20 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN DUPLICATE 11

ACCESSION NUMBER: 1998:302670 BIOSIS
DOCUMENT NUMBER: PREV199800302670
TITLE: Homology model for the ligand-binding domain of the human estrogen receptor.
AUTHOR(S): Maaloluf, George J.; Xu, Wenrog; Smith, Temple F.; Mohr, Scott C. (1)
CORPORATE SOURCE: (1) Boston Univ., Dep. Chem., 590 Commonwealth Ave., Boston, MA 02215 USA
SOURCE: Journal of Biomolecular Structure and Dynamics, (April, 1998) Vol. 15, No. 5, pp. 841-851.
ISSN: 0739-1102.
DOCUMENT TYPE: Article
LANGUAGE: English

AB We have modeled the ligand-binding domain (LBD) of the human estrogen receptor **protein** (hER) by homology to the known crystal **structure** of the LBD of the a isoform of human retinoate-X receptor (hRX). **Alignment** of hER with members of the nuclear receptor superfamily defined probable secondary structures which we used to constrain backbone torsion angles and hydrogen bonds. From published studies we identified key interactions between hER and estradiol to use to

dock the hormone in its ligand-binding pocket. Since the hRX crystal **structure** corresponds to the unliganded form of the LBD, we adopted the "mousetrap" mechanism proposed by Renaud et al. to predict the **structure** of the E2-bound hER. Refinement by molecular dynamics and **energy minimization** gave a model which matches well the known facts about the estradiol phamacophore. It also provides a possible explanation for how hER discriminates between estradiol and testosterone.

=> d ibib ab l29 21-44

L29 ANSWER 21 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
DUPLICATE 12

ACCESSION NUMBER: 1997:125238 BIOSIS
DOCUMENT NUMBER: PREV199799431741
TITLE: Three-dimensional **structure** of toxin OSK1 from Orthochirus scrobiculosus scorpion venom.
AUTHOR(S): Jaravine, Victor A.; Nolde, Dmitry E.; Reibarkh, Michail J.; Korolkova, Yulia V.; Kozlov, Sergey A.; Pluzhnikov, Kirill A.; Grishin, Eugene V.; Arseniev, Alexander S. (1)
CORPORATE SOURCE: (1) Shemyakin and Ovchinnikov Inst. Bioorganic Chem., Russian Acad. Sci., Ul. Miklukho-Maklaya 16/10, Moscow 117871 Russia
SOURCE: Biochemistry, (1997) Vol. 36, No. 6, pp. 1223-1232. ISSN: 0006-2960.
DOCUMENT TYPE: Article
LANGUAGE: English

AB A 600 MHz 1H NMR study of toxin OSK1, blocker of small-conductance Ca-2+-activated K+ channels, is presented. The unambiguous sequential assignment of all the protons of the toxin was obtained using TOCSY, DQF-COSY, and NOESY experiments at pH 3.0 (10, 30, and 45 degree C) in aqueous solution. 3J-Nalpha, 3J-alpha-beta vicinal spin coupling constants were determined in high-resolution spectra. The cross-peak volumes in NOESY spectra and the coupling constants were used to define the local **structure** of the **protein** by the program HABAS and to generate torsion angle and interproton distance constraints for the program DIANA. Hydrogen-deuterium exchange rates of amide protons showed possible locations of hydrogen bonds. The hydrogen bond acceptors and disulfide bridges between residues 8-28, 14-33, and 18-35 were determined when analyzing distance distribution in preliminary DIANA structures. All constraints were used to obtain a set of 30 structures by DIANA. The resulting rms deviations over 30 structures are 1.30 ANG for the heavy atoms and 0.42 ANG for the backbone heavy atoms. The structures were refined by constrained **energy minimization** using the SYBYL program. Their analysis indicated the existence of the alpha-helix (residues 10-21) slightly distorted at the Cys 14 residue, two main strands of the antiparallel beta-sheet (24-29, 32-38). and the extended fragment (2-6). The motif is stabilized by the disulfide bridges in the way common to all known scorpion toxins. Using the fine spatial toxin **structure**, **alignment** of the homologues, mutagenesis analysis, and comparison of scorpion toxin family functions, we delineate some differences significant for the toxin specificity.

L29 ANSWER 22 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
DUPLICATE 13

ACCESSION NUMBER: 1998:88289 BIOSIS
DOCUMENT NUMBER: PREV199800088289
TITLE: Comparative modeling and molecular dynamics studies of the delta, kappa and mu opioid receptors.
AUTHOR(S): Strahs, Daniel; Weinstein, Harel (1)
CORPORATE SOURCE: (1) Dep. Physiol. Biophysics, Mount Sinai Sch. Med., New York, NY 10029-6574 USA
SOURCE: Protein Engineering, (Sept., 1997) Vol. 10, No. 9, pp.

1019-1038.
ISSN: 0269-2139.
DOCUMENT TYPE: Article
LANGUAGE: English

AB Molecular models of the trans-membrane domains of delta, kappa and mu opioid receptors, members of the G-**protein** coupled receptor (GPCR) superfamily, were developed using techniques of homology modeling and molecular dynamics simulations. Structural elements were predicted from sequence **alignments** of opioid and related receptors based on (i) the consensus, periodicities and biophysical interpretations of **alignment**-derived properties, and (ii) tertiary **structure** homology to rhodopsin. Initial model structures of the three receptors were refined computationally with **energy minimization** and the result of the first 210 ps of a 2 ns molecular dynamics trajectory at 300K. Average structures from the trajectory obtained for each receptor subtype after release of the initial backbone constraints show small backbone deviations, indicating stability. During the molecular dynamics phase, subtype-differentiated residues of the receptors developed divergent structures within the models, including changes in regions common to the three subtypes and presumed to belong to ligand binding regions. The divergent features developed by the model structures appear to be consistent with the observed ligand binding selectivities of the opioid receptors. The results thus implicate identifiable receptor microenvironments as primary determinants of some of the observed subtype specificities in opiate ligand binding and in functional effects of mutagenesis. Networks of interacting residues observed in the models are common to the opiate receptors and other GPCRs, indicating core interfaces that are potentially responsible for structural integrity and signal transduction. Analysis of extended molecular dynamics trajectories reveals concerted motions of distant parts of ligand-binding regions, suggesting motion-sensitive components of ligand binding. The comparative modeling results from this study help clarify experimental observations of subtype differences and suggest both structural and dynamic rationales for differences in receptor properties.

L29 ANSWER 23 OF 44 BIOSIS
DUPLICATE 14
ACCESSION NUMBER: 1997:308245 BIOSIS
DOCUMENT NUMBER: PREV199799616048
TITLE: Ligand-based **protein alignment** and isozyme specificity of glutathione S-transferase inhibitors.
AUTHOR(S): Koehler, Ryan T.; Villar, Hugo O. (1); Bauer, Karin E.; Higgins, Deborah L.
CORPORATE SOURCE: (1) Chem. Dep. Terrapin Technologies, 750 Gateway Blvd., South San Francisco, CA 94080 USA
SOURCE: Proteins Structure Function and Genetics, (1997) Vol. 28, No. 2, pp. 202-216.
ISSN: 0887-3585.
DOCUMENT TYPE: Article
LANGUAGE: English

AB Glutathione S-transferases (GST, E.C.2.5.1.18) comprise a family of detoxification enzymes. Elevated levels of specific GST isozymes in tumor cells are thought responsible for resistance to chemotherapeutics, which renders selective GST inhibitors potentially useful pharmaceutical agents. We discuss the development of a **structure** activity model that rationalizes the isozyme selectivity observed in a series of 12 glutathione (GSH) analogues. Enzymatic activity data was determined for human P1-1, A1-1, and M2-2 isozymes, and these data were then considered in light of structural features of these three GST proteins. A survey of all GST structures in the PDB revealed that GSH binds to these proteins in a single "bioactive" conformation. To focus on differences between binding sites, we exploited our finding of a common GSH conformation and **aligned** the GST x-ray structures using bound ligands rather than

the backbones of the different proteins. Once **aligned**, binding site lipophilicity and electrostatic potentials were computed, visualized, and compared. Docking and **energy minimization** exercises provided additional refinements to a model of selectivity developed initially by visual analysis. Our results suggest that binding site shape and lipophilic character are key determinants of GST isozyme selectivity for close GSH analogues.

L29 ANSWER 24 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
DUPLICATE 15

ACCESSION NUMBER: 1998:130564 BIOSIS
DOCUMENT NUMBER: PREV199800130564
TITLE: Model building by comparison: A combination of expert knowledge and computer automation.
AUTHOR(S): Bates, Paul A.; Jackson, Richard M.; Sternberg, Michael J. E. (1)
CORPORATE SOURCE: (1) Biomolecular Modelling Lab., Imperial Cancer Res. Fund, Lincoln's Inn Fields, P.O. Box 123, London WC2A 3PX UK
SOURCE: Proteins, (1997) Vol. 0, No. SUPPL. 1, pp. 59-67. ISSN: 0887-3585.
DOCUMENT TYPE: Article
LANGUAGE: English

AB The CASP blind trials (Critical Assessment of techniques for **protein Structure** Prediction) assess the accuracy of **protein** prediction that includes evaluation of comparative model building of **protein** structures. Comparative models of four proteins (T0001, T0003, T0017, and T0028) for CASP2 (held during 1996) were constructed using computer algorithms combined with visual inspection. Essentially the main-chain modelling involves construction of the target **structure** from rigid-body segments of homologues and loop fragments extracted from homologous and nonredundant databases. Side-chains were initially constructed by inheritance from the parent or from a rotamer library. Side-chain conformations were then refined using a novel mean field approach that includes solvation. Comparison of the models with the subsequently released X-ray structures identified the successes and limitations of our approach. The most problematic area is the quality of the sequence **alignments** between parent(s) and target. In this respect the overinterpretation of the conserved features within homologous families can be misleading. Several features of our approach have a positive effect on the accuracy of the models. For T0003, inspection correctly identified that a lower sequence identity parent provides the best framework for this model. Loop selection worked well where a homologous **protein** fragment was used, but that the use of nonredundant fragment library remains problematic for hinge movements and displacements in secondary **structure** elements relative to the parent. Side-chain refinement improved residue conformations relative to the initial model. Use of limited **energy minimization** improved the stereochemical quality of the model without increasing the RMS deviation. This study has identified methods that are effective and areas requiring further attention to improve model building by comparison.

L29 ANSWER 25 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
DUPLICATE 16

ACCESSION NUMBER: 1996:519501 BIOSIS
DOCUMENT NUMBER: PREV199699241857
TITLE: Modeling of the D1/D2 proteins and cofactors of the photosystem II reaction center: Implications for herbicide and bicarbonate binding.
AUTHOR(S): Xiong, Jin; Subramaniam, Shankar; Govindjee (1)
CORPORATE SOURCE: (1) Dep. Plant Biol., 265 Morrill Hall, 505 S. Goodwin Ave., Univ. Ill. at Urbana-Champaign, Urbana, IL 61801-3707 USA
SOURCE: Protein Science, (1996) Vol. 5, No. 10, pp. 2054-2073. ISSN: 0961-8368.

DOCUMENT TYPE: Article
LANGUAGE: English

AB A three-dimensional model of the photosystem II (PSII) reaction center from the cyanobacterium Synechocystis sp. PCC 6803 was generated based on homology with the anoxygenic purple bacterial photosynthetic reaction centers of Rhodobacter sphaeroides and Rhodopseudomonas viridis, for which the X-ray crystallographic structures are available. The model was constructed with an **alignment** of D1 and D2 sequences with the L and M subunits of the bacterial reaction center, respectively, and by using as a scaffold the structurally conserved regions (SCRs) from bacterial templates. The structurally variant regions were built using a novel sequence-specific approach of searching for the best-matched **protein** segments in the **Protein** Data Bank with the "basic local **alignment** search tool" (Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ, 1990, J Mol Biol 215:403-410), and imposing the matching conformational preference on the corresponding D1 and D2 regions. The **structure** thus obtained was refined by **energy minimization**. The modeled D1 and D2 proteins contain five transmembrane a-helices each, with cofactors (4 chlorophylls, 2 pheophytins, 2 plastoquinones, and a non-heme iron) essential for PSII primary photochemistry embedded in them. A beta-carotene, considered important for PSII photoprotection, was also included in the model. Four different possible conformations of the primary electron donor P680 chlorophylls were proposed, one based on the homology with the bacterial template and the other three on existing experimental suggestions in literature. The P680 conformation based on homology was preferred because it has the lowest energy. Redox active tyrosine residues important for P680+ reduction as well as residues important for PSII cofactor binding were analyzed. Residues involved in interprotein interactions in the model were also identified. Herbicide 3-(3,4-dichlorophenyl)-1,1-dimethylurea (DCMU) was also modeled in the plastoquinone QB binding niche using the structural information available from a DCMU-binding bacterial reaction center. A bicarbonate anion, known to play a role in PSII, but not in anoxygenic photosynthetic bacteria, was modeled in the non-heme iron site, providing a bidentate ligand to the iron. By modifying the previous hypothesis of Blubaugh and Govindjee (1988, Photosyn Res 19:85-128), we modeled a second bicarbonate and a water molecule in the Q-B site and we proposed a hypothesis to explain the mechanism of Q-B protonation mediated by bicarbonate and water. The bicarbonate, stabilized by D1-R257, donates a proton to Q-B-2- through the intermediate of D1-H252; and a water molecule donates another proton to Q-B-2-. Based on the discovery of a "water transport channel" in the bacterial reaction center, an analogous channel for transporting water and bicarbonate is proposed in our PSII model. The putative channel appears to be primarily positively charged near QB and the non-heme iron, in contrast to the polarity distribution in the bacterial water transport channel. The constructed model has been found to be consistent with most existing data.

L29 ANSWER 26 OF 44 BIOSIS
DUPLICATE 17

ACCESSION NUMBER: 1996:282774 BIOSIS
DOCUMENT NUMBER: PREV199699005130

TITLE: Thermodynamic prediction of conserved secondary **structure**: Application to the RRE element of HIV, the tRNA-like element of CMV and the mRNA of prion protein.

AUTHOR(S): Lueck, Rupert; Steger, Gerhard; Riesner, Detlev (1)
CORPORATE SOURCE: (1) Biologische-Med. Forschungzentrum, Heinrich-Heine-Univ., Duesseldorf Germany

SOURCE: Journal of Molecular Biology, (1996) Vol. 258, No. 5, pp. 813-826.
ISSN: 0022-2836.

DOCUMENT TYPE: Article
LANGUAGE: English

AB An algorithm for prediction of conserved secondary **structure** of

single-stranded RNA is presented. For each RNA of a set of homologous RNAs optimal and suboptimal secondary structures are calculated and stored in a base-pair probability matrix. A multiple sequence **alignment** is performed for the set of RNAs. The resulting gaps are introduced into the individual probability matrices. These homologous probability matrices are summed to give a consensus probability matrix emphasizing the conserved secondary **structure** elements of the RNA set. Thus the algorithm combines the advantages of thermodynamic **structure** prediction by **energy minimization** with the information obtained from phylogenetic **alignment** of sequences. The algorithm is applied to three examples. The REV-responsive element of HIV, the **structure** of which is well known from the literature, was chosen to test the algorithm. The second example is the 3' terminal segment of genomic single-stranded RNAs of cucumber mosaic viruses; a **structure** similar to that of the related brome mosaic virus was expected and was confirmed. The third example is the prion-**protein** mRNA from different organisms; the **structure** of this mRNA is not known. By application of the algorithm highly conserved hairpins were found in the prion-**protein** mRNA.

L29 ANSWER 27 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
ACCESSION NUMBER: 1996:485590 BIOSIS
DOCUMENT NUMBER: PREV199699200846
TITLE: Modeling spatial **structure** of obelin, a hydroid calcium-activated photoprotein.
AUTHOR(S): Sandalova, T. P.
CORPORATE SOURCE: Inst. Biophys., Sib. Div., Russ. Acad. Sci., Krasnoyarsk 660036 Russia
SOURCE: Molekulyarnaya Biologiya (Moscow), (1996) Vol. 30, No. 3, pp. 621-630.
ISSN: 0026-8984.
DOCUMENT TYPE: Article
LANGUAGE: Russian
SUMMARY LANGUAGE: Russian
AB A model of three-dimensional **structure** was proposed for obelin a photoprotein from Obelia longissima. Amino acid sequence of sarcoplasmic calcium-binding invertebrate **protein** was transformed into obelin sequence according to **alignment**: then the **energy minimization** of the obtained **protein** was performed. The analysis of the model showed that the latter was a compact globule with pronounced hydrophobic nucleus and satisfactory stereochemistry, that is., possessed all properties of a globular **protein**; the model contained a cavity lined with residues affecting photoprotein activity. The size of the cavity was sufficient for binding a cofactor. Therefore, it was assumed that this cavity was the active center of photoproteins. Based on the results of the spatial **structure** of the model, it was proposed to use several obelin residues for mutation experiments.

L29 ANSWER 28 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
DUPLICATE 18
ACCESSION NUMBER: 1996:150118 BIOSIS
DOCUMENT NUMBER: PREV199698722253
TITLE: Prediction of the three-dimensional **structure** of the rap-1A protein from its homology to the ras-gene-encoded p21 protein.
AUTHOR(S): Chen, James M. (1); Grad, Rosalyn; Monaco, Regina; Pincus, Matthew R.
CORPORATE SOURCE: (1) Dupont Agric. Products, Stein Haskell Res. Cent., PO Box 30, Newark, DE 19714 USA
SOURCE: Journal of Protein Chemistry, (1996) Vol. 15, No. 1, pp. 11-16.
ISSN: 0277-8033.
DOCUMENT TYPE: Article
LANGUAGE: English

AB rap-1A, an anti-oncogene-encoded **protein**, is a ras-p21-like **protein** whose sequence is over 80% homologous to p21 and which interacts with the same intracellular target proteins and is activated by the same mechanisms as p21, e.g., by binding GTP in place of GDP. Both interact with effector proteins in the same region, involving residues 32-47. However, activated rap-1A blocks the mitogenic signal transducing effects of p21. Optimal sequence **alignment** of p21 and rap-1A shows two insertions of rap-1A at ras positions 120 and 138. We have constructed the three-dimensional **structure** of rap-1A bound to GTP by using the **energy-minimized** three-dimensional **structure** of ras-p21 as the basis for the modeling using a stepwise procedure in which identical and homologous amino acid residues in rap-1A are assumed to adopt the same conformation as the corresponding residues in p21. Side-chain conformations for homologous and nonhomologous residues are generated in conformations that are as close as possible to those of the corresponding side chains in p21. The entire **structure** has been subjected to a nested series of **energy minimizations**. The final predicted **structure** has an overall backbone deviation of 0.7 ANG from that of ras-p21. The effector binding domains from residues 32-47 are identical in both proteins (except for different side chains of different residues at position 45). A major difference occurs in the insertion region at residue 120. This region is in the middle of another effector loop of the p21 **protein** involving residues 115-126. Differences in sequence and **structure** in this region may contribute to the differences in cellular functions of these two proteins.

L29 ANSWER 29 OF 44 EMBASE

ACCESSION NUMBER: 95246528 EMBASE
DOCUMENT NUMBER: 1995246528
TITLE: Asymmetry in the **structure** of glycopeptide antibiotic dimers: NMR studies of the ristocetin A complex with a bacterial cell wall analogue.
AUTHOR: Groves P.; Searle M.S.; Waltho J.P.; Williams D.H.
CORPORATE SOURCE: Center for Molecular Recognition, University Chemical Laboratories, University of Cambridge, Lensfield Road,Cambridge CB2 1EW, United Kingdom
SOURCE: Journal of the American Chemical Society, (1995) 117/30 (7958-7964).
ISSN: 0002-7863 CODEN: JACSAT
COUNTRY: United States
DOCUMENT TYPE: Journal; Article
FILE SEGMENT: 004 Microbiology
037 Drug Literature Index
LANGUAGE: English
SUMMARY LANGUAGE: English

AB The solution **structure** of a dimer complex of the **glycopeptide** antibiotic ristocetin A has been determined from NOE constraints, **energy minimization**, and molecular dynamics calculations. The **structure** is that of an asymmetric dimer in which the conformation of the two monomeric units differs in the orientation of the tetrasaccharide attached to the aromatic ring of residue 4. Although hydrogen bonding interactions between the **peptide** backbones of the two antibiotic monomers occur in a symmetrical head-to-tail orientation, the overall dimer assymmetry arises as a consequence of a parallel, head-to-head **alignment** of the tetrasaccharides. Thus, in the two monomeric antibiotic conformations that constitute the dimer, the orientations of the tetrasaccharides are related by an .apprx.180.degree. rotation about the glucose-ring 4 glycosidic bond. The quite different orientation of the tetrasaccharide in each half of the dimer results in significant differences in binding interactions with cell wall peptides occupying the two different sites on the dimer. In one site, the hydrophobic face of glucose interacts with the methyl group of the C-terminal D-alanine of cell wall analogues, while the rhamnose

sugar of the same tetrasaccharide may act as a hydrophilic 'cap' where three hydroxyl groups on the edge of the sugar can mimic a group of water molecules through a network of hydrogen bonds. An arabinose sugar of the other tetrasaccharide occupies a similar position to the rhamnose in the second ligand binding site; its single hydroxyl group may be less effective as a hydrophilic cap, and the hydrophobic interaction to a glucose face (see above) cannot now take place. These observations lead to the conclusion that there may be a marked difference in the ligand binding affinities for the two sites. This conclusion has been confirmed experimentally.

L29 ANSWER 30 OF 44 BIOSIS
DUPLICATE 19

ACCESSION NUMBER: 1995:499434 BIOSIS
DOCUMENT NUMBER: PREV199598522984
TITLE: Predicting the **structure** of the light-harvesting complex II of Rhodospirillum molischianum.
AUTHOR(S): Hu, Xiche; Xu, Dong; Hamer, Kenneth; Schulten, Kloaus (1); Koepke, Juergen; Michel, Hartmut
CORPORATE SOURCE: (1) Theoretical Biophysics, Beckman Inst., 405 North Mathews Ave., Urbana, IL 61801 USA
SOURCE: Protein Science, (1995) Vol. 4, No. 9, pp. 1670-1682.
ISSN: 0961-8368.
DOCUMENT TYPE: Article
LANGUAGE: English

AB We attempted to predict through computer modeling the **structure** of the light-harvesting complex II (LH-II) of Rhodospirillum molischianum, before the impending publication of the **structure** of a homologous **protein** solved by means of X-ray diffraction. The **protein** studied is an integral membrane **protein** of 16 independent polypeptides, 8 alpha-apoproteins and 8 beta-apoproteins, which aggregate and bind to 24 bacteriochlorophyll-a's and 12 lycopenes. Available diffraction data of a crystal of the **protein**, which could not be phased due to a lack of heavy metal derivatives, served to test the predicted **structure**, guiding the search. In order to determine the secondary **structure**, hydropathy analysis was performed to identify the putative transmembrane segments and multiple sequence **alignment** propensity analyses were used to pinpoint the exact sites of the 20-residue-long transmembrane segment and the 4-residue-long terminal sequence at both ends, which were independently verified and improved by homology modeling. A consensus assignment for the secondary **structure** was derived from a combination of all the prediction methods used. Three-dimensional structures for the alpha- and the beta-apoprotein were built by comparative modeling. The resulting tertiary structures are combined, using X-PLOR, into an alpha-beta dimer pair with bacteriochlorophyll-a's attached under constraints provided by site-directed mutagenesis and spectral data. The alpha-beta dimer pairs were then aggregated into a quaternary **structure** through further molecular dynamics simulations and **energy minimization**. The **structure** of LH-II so determined is an octamer of alpha-beta heterodimers forming a ring with a diameter of 70 ANG .

L29 ANSWER 31 OF 44 BIOSIS
DUPLICATE 20

ACCESSION NUMBER: 1995:215525 BIOSIS
DOCUMENT NUMBER: PREV199598229825
TITLE: Comparative modeling of the three-dimensional **structure** of Type II antifreeze protein.
AUTHOR(S): Sonnichsen, Frank D. (1); Sykes, Brian D.; Davies, Peter L.
CORPORATE SOURCE: (1) Protein Eng. Network Cent. Excellence, Dep. Biochem., Heritage Med. Res. Cent. 7-13, Univ. Alberta, Edmonton, AB T6G 2S2 Canada
SOURCE: Protein Science, (1995) Vol. 4, No. 3, pp. 460-471.
ISSN: 0961-8368.

DOCUMENT TYPE: Article
LANGUAGE: English

AB Type II antifreeze proteins (AFP), which inhibit the growth of seed ice crystals in the blood of certain fishes (sea raven, herring, and smelt), are the largest known fish AFPs and the only class for which detailed structural information is not yet available. However, a sequence homology has been recognized between these proteins and the carbohydrate recognition domain of C-type lectins. The **structure** of this domain from rat mannose-binding **protein** (MBP-A) has been solved by X-ray crystallography (Weis WI, Drickamer K, Hendrickson WA, 1992, Nature 360:127-134) and provided the coordinates for constructing the three-dimensional model of the 129-amino acid Type 11 AFP from sea raven, to which it shows 19% sequence identity. Multiple sequence **alignments** between Type 11 AFPs, pancreatic stone **protein** , MBP-A, and as many as 50 carbohydrate-recognition domain sequences from various lectins were performed to determine reliably **aligned** sequence regions. Successive molecular dynamics and **energy minimization** calculations were used to relax bond lengths and angles and to identify flexible regions. The derived **structure** contains two alpha-helices, two beta-sheets, and a high proportion of amino acids in loops and turns. The model is in good agreement with preliminary NMR spectroscopic analyses. It explains the observed differences in calcium binding between sea raven Type II AFP and MBP-A. Furthermore, the model proposes the formation of five disulfide bridges between Cys 7 and Cys 18, Cys 35 and Cys 125, Cys 69 and Cys 100, Cys 89 and Cys III, and Cys 101 and Cys 117. Based on the predicted features of this model, a site for **protein**-ice interaction is proposed.

L29 ANSWER 32 OF 44 BIOSIS
DUPLICATE 21

ACCESSION NUMBER: 1996:31265 BIOSIS
DOCUMENT NUMBER: PREV199698603400
TITLE: A critical assessment of comparative molecular modeling of tertiary structures of proteins.
AUTHOR(S): Mosimann, Steven; Mleshko, Ron; James, Michael N. G. (1)
CORPORATE SOURCE: (1) Med. Res. Council Canada, Group Protein Structure Function, Dep. Biochem., Univ. Alberta, Edmonton, AB T6G 2H7 Canada
SOURCE: Proteins Structure Function and Genetics, (1995) Vol. 23, No. 3, pp. 301-317.
ISSN: 0887-3585.
DOCUMENT TYPE: Article
LANGUAGE: English

AB In spite of the tremendous increase in the rate at which **protein** structures are being determined, there is still an enormous gap between the numbers of known DNA-derived sequences and the numbers of three-dimensional structures. In order to shed light on the biological functions of the molecules, researchers often resort to comparative molecular modeling. Earlier work has shown that when the sequence **alignment** is in error, then the comparative model is guaranteed to be wrong. In addition, loops, the sites of insertions and deletions in families of homologous proteins, are exceedingly difficult to model. Thus, many of the current problems in comparative molecular modeling are minor versions of the global **protein** folding problem. In order to assess objectively the current state of comparative molecular modeling, 13 groups submitted blind predictions of seven different proteins of undisclosed tertiary **structure**. This assessment shows that where sequence identity between the target and the template **structure** is high (gt 70%), comparative molecular modeling is highly successful. On the other hand, automated modeling techniques and sophisticated **energy minimization** methods fail to improve upon the starting structures when the sequence identity is low (apprx 30%). Based on these results it appears that insertions and deletions are still major problems. Successfully deducing the correct sequence **alignment**

when the local similarity is low is still difficult. We suggest some minimal testing of submitted coordinates that should be required of authors before papers on comparative molecular modeling are accepted for publication in journals.

L29 ANSWER 33 OF 44 BIOSIS
DUPLICATE 22

ACCESSION NUMBER: 1995:304492 BIOSIS
DOCUMENT NUMBER: PREV199598318792
TITLE: Purification, cloning, and sequencing of archaebacterial pyrophosphatase from the extreme thermoacidophile Sulfolobus acidocaldarius.
AUTHOR(S): Meyer, W.; Moll, R.; Kath, T.; Schaefer, G. (1)
CORPORATE SOURCE: (1) Inst. Biochem., Med. Univ., Ratzeburger Allee 160, 23538 Lubeck Germany
SOURCE: Archives of Biochemistry and Biophysics, (1995) Vol. 319, No. 1, pp. 149-156.
ISSN: 0003-9861.
DOCUMENT TYPE: Article
LANGUAGE: English

AB Cytoplasmic pyrophosphatases are indispensible for the function of cellular bioenergetics. From the extreme thermoacidophilic archaeon Sulfolobus acidocaldarius, situated at one of the lowest branches of the phylogenetic tree, a cytosolic pyrophosphatase has been isolated and purified 200-fold to electrophoretic homogeneity by combining ion-exchange and gel-exclusion chromatography. The native enzyme consists of a homotetramer of 71 kDa apparent molecular mass; the subunit displays an apparent molecular mass of 17 kDa on sodium dodecyl sulfate-polyacrylamide gel electrophoresis. The enzyme has an absolute requirement for divalent cations (Mg-2+) and a temperature optimum of 75 degree C coinciding with the growth optimum of the organism; the apparent estimated activation energy is 79.5 kJ/mol. A large variety of cytosolic extracts from other archaebacteria has been probed with a polyclonal antiserum raised against the purified **protein**; surprisingly, except for an extremely weak signal with S. solfataricus none of the other organisms showed any cross-reactivity. Also, Escherichia coli PPase does not cross-react. Based on N-terminal sequencing the gene has been cloned and sequenced. It codes for a 173-amino-acid **protein** with a calculated molecular mass of 19,365 kDa. **Alignment** with known eucaryotic and procaryotic PPases reveals invariant conservation of all residues presently assumed to be involved in metal and substrate binding. Unexpectedly, the highest similarity is found with the enzyme from the phylogenetically extremely distant eubacterium E. coli, but immunological cross-reactivity is absent. Similarity to the only known other archaebacterial PPase is much weaker. Using the 3D **structure** of the Thermus thermophilus enzyme as a scaffold an **energy-minimized** structural model is presented, deviating only minimally from the former. The structural features are discussed. The enzyme provides an excellent model for studies of thermostability and folding dynamics since heterologous overexpression has been achieved and genetically mutated forms become accessible.

L29 ANSWER 34 OF 44 BIOSIS
DUPLICATE 23

ACCESSION NUMBER: 1994:430601 BIOSIS
DOCUMENT NUMBER: PREV199497443601
TITLE: Comparative modelling of major house dust mite allergen Der p I: **Structure** validation using an extended environmental amino acid propensity table.
AUTHOR(S): Topham, Christopher M. (1); Srinivasan, N.; Thorpe, Christopher J.; Overington, John P.; Kalsheker, Noor A.
CORPORATE SOURCE: (1) St. Luke's Inst. Cancer Res., Dep. Chem., Univ. Coll. Dublin, Belfied Ireland
SOURCE: Protein Engineering, (1994) Vol. 7, No. 7, pp. 869-894.
ISSN: 0269-2139.

DOCUMENT TYPE: Article
LANGUAGE: English

AB A model of the 3-D **structure** of a major house dust mite allergen Der p I associated with hypersensitivity reactions in humans was built from its amino acid sequence and its homology to three known structures, papain, actinidin and papaya proteinase O of the cysteine proteinase family. Comparative modelling using COMPOSER was used to arrive at an initial model. This was refined using interactive graphics and **energy minimization** with the AMBER force field incorporated in SYBYL (Tripos Associates). Compatibility of the Der p I amino acid sequence with the cysteine proteinase fold was checked using an environment-dependent amino acid propensity table incorporated into a new program HARMONY with a variable length windowing facility. A five residue window was used to probe local conformational integrity. Propensities were derived from a structural **alignment** database of homologous proteins using a robust entropy-driven smoothing procedure. Der p I shares essential structural and mechanistic features with other papain-like cysteine proteinases, including cathepsin B. The active-site thiolate-imidazolium ion pair comprises the side chains of Cys34 and His170. A cystine disulfide not present in other known **structure** bridges residue 4 of an N-terminal extension and the core residue 117. Two conserved disulfide bridges are formed by residues 31 and 71 and residues 65 and 103. Model building of **peptide** substrate analogue complexes suggests a preference for phenylalanyl or bask residues at the P-2 position, whilst selectivity may be of minor importance at the S-1 subsite. The electrostatic influences on the Der p I active-site ion pair and extended **peptide** binding region are markedly different from those in known structures. A highly immunogenic surface exposed region (residues 107-131), comprising several overlapping T cell epitope sites, has no shared sequence identity with human liver cathepsin B and contains three insertion - deletion sites. The **structure** provides a basis for testing the substrate specificity of Der p I and the potential role of proteinase activity in hypersensitivity reactions. These studies may offer a new treatment strategy by hyposensitization with inactive mutants or mutants with significantly altered proteinase activity, either alone or complexed with antibody.

L29 ANSWER 35 OF 44 BIOSIS
DUPLICATE 24

ACCESSION NUMBER: 1994:35947 BIOSIS
DOCUMENT NUMBER: PREV199497048947
TITLE: Myelin P-0-glycoprotein: Predicted **structure** and interactions of extracellular domain.
AUTHOR(S): Wells, Christopher A.; Saavedra, Raul A.; Inouye, Hideyo; Kirschner, Daniel A. (1)
CORPORATE SOURCE: (1) Neurology Res. Enders 2, Children's Hosp., 300 Longwood Ave., Boston, MA 02115 USA
SOURCE: Journal of Neurochemistry, (1993) Vol. 61, No. 6, pp. 1987-1995.
ISSN: 0022-3042.
DOCUMENT TYPE: Article
LANGUAGE: English

AB **Protein** zero (P-0), a transmembrane glycoprotein, accounts for over 50% of the total **protein** in PNS myelin. The extracellular domain of P-0 (P-0-ED) is similar to the immunoglobulin variable domain, carrying one acceptor sequence for N-linked glycosylation. The x-ray diffraction analysis of PNS myelin has demonstrated reversible transitions that depend on pH and ionic strength, resulting in three distinct structures characterized by widths of about 36 ANG , 50 ANG (native), and 90 ANG between the extracellular surfaces of the membranes. In the current work, we considered the constraints imposed by these x-ray diffraction data on the orientation of P-0-ED, and we propose how this immunoglobulin-like domain could be accommodated in the variable widths of the extracellular space between myelin membranes. The modeling made use of

the finding that beta-strand predictions for P-0-ED are virtually superimposable with those of the V-H domain of the phosphocholine-binding immunoglobulin M603 of mouse, which has a similar number of residues as P-0-ED and a **structure** that has been solved crystallographically. The dimensions of P-0-ED from the space-filling model, developed using PC-based molecular modeling software, were found to be 44 ANG times 25 ANG times 23 ANG . On the assumption that neither the shape nor the orientation of P-0-ED changes appreciably, then the different widths at the extracellular apposition would easily accommodate P-0-ED from apposed membranes if the molecules were oriented so that the beta-strands were approximately perpendicular to the membrane surface. The apposed P-0-EDs would fully overlap at the closest apposition of the membranes, partially overlap in the native state, and **align** end to end in the incompletely swollen state. The P-0-ED regions analogous to the complementarity-determining regions of immunoglobulins can account for the recognition of P-0-ED from apposed membranes in the incompletely swollen state. Two of the faces of P-0-ED that show charge complementarity could account for the homophilic interactions of P-0-ED from apposed membranes in the native state. This association can be stabilized further by hydrophobic interactions. The N-linked nonasaccharide after **energy minimization** fit into a cavity, which was at the base of P-0-ED and which was lined with three positively charged residues. Thus, the carbohydrate may help to maintain the orientation of P-0 at the membrane surface. Our model shows how the single immunoglobulin-like domain of P-0 can account for distinct structural states of myelin membrane packing by homophilic interactions.

L29 ANSWER 36 OF 44 MEDLINE on STN DUPLICATE 25
ACCESSION NUMBER: 94073084 MEDLINE
DOCUMENT NUMBER: 94073084 PubMed ID: 7504550
TITLE: Calculation of **protein** backbone geometry from alpha-carbon coordinates based on **peptide**-group dipole **alignment**.
AUTHOR: Liwo A; Pincus M R; Wawak R J; Rackovsky S; Scheraga H A
CORPORATE SOURCE: Baker Laboratory of Chemistry, Cornell University, Ithaca, New York 14853-1301.
CONTRACT NUMBER: CA 42500 (NCI)
GM-14312 (NIGMS)
SOURCE: PROTEIN SCIENCE, (1993 Oct) 2 (10) 1697-714.
Journal code: 9211750. ISSN: 0961-8368.
PUB. COUNTRY: United States
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 199401
ENTRY DATE: Entered STN: 19940203
Last Updated on STN: 19960129
Entered Medline: 19940113

AB An algorithm is proposed for the conversion of a virtual-bond polypeptide chain (connected C alpha atoms) to an all-atom backbone, based on determining the most extensive hydrogen-bond network between the peptide groups of the backbone, while maintaining all of the backbone atoms in energetically feasible conformations. Hydrogen bonding is represented by **aligning** the **peptide**-group dipoles. These peptide groups are not contiguous in the amino acid sequence. The first dipoles to be aligned are those that are both sufficiently close in space to be arranged in approximately linear arrays termed dipole paths. The criteria used in the construction of dipole paths are: to assure good alignment of the greatest possible number of dipoles that are close in space; to optimize the electrostatic interactions between the dipoles that belong to different paths close in space; and to avoid locally unfavorable amino acid residue conformations. The equations for dipole alignment are solved separately for each path, and then the remaining single dipoles are aligned optimally with the electrostatic field from the dipoles that

belong to the dipole-path network. A least-squares minimizer is used to keep the geometry of the alpha-carbon trace of the resulting backbone close to that of the input virtual-bond chain. This procedure is sufficient to convert the virtual-bond chain to a real chain; in applications to real systems, however, the final **structure** is obtained by minimizing the total ECEPP/2 (empirical conformational energy program for peptides) energy of the system, starting from the geometry resulting from the solution of the **alignment** equations. When applied to model alpha-helical and beta-sheet structures, the algorithm, followed by the ECEPP/2 **energy minimization**, resulted in an energy and backbone geometry characteristic of these alpha-helical and beta-sheet structures. Application to the alpha-carbon trace of the backbone of the crystallographic 5PTI **structure** of bovine pancreatic trypsin inhibitor, followed by ECEPP/2 **energy minimization** with C alpha-distance constraints, led to a **structure** with almost as low energy and root mean square deviation as the ECEPP/2 geometry analog of 5PTI, the best agreement between the crystal and reconstructed backbone being observed for the residues involved in the dipole-path network.

L29 ANSWER 37 OF 44 BIOTECHDS COPYRIGHT 2003 THOMSON DERWENT/ISI on STN
ACCESSION NUMBER: 1993-14710 BIOTECHDS
TITLE: Molecular cloning of a hyperthermophilic archaebacterial adenylate-kinase;
Sulfolobus acidocaldarius enzyme characterization (conference abstract)
AUTHOR: Kath T; Schaefer G
LOCATION: Institut fuer Biochemie, Med. Universitaet zu Luebeck, Ratzeburger Allee 160, 23538 Luebeck, Germany.
SOURCE: Biol.Chem.Hoppe Seyler; (1993) 374, 9, 770-71
CODEN: BCHSEI
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Adenylate-kinase (AK, EC-2.7.4.3) was purified to homogeneity from the thermoacidophilic archaeon Sulfolobus acidocaldarius and characterized. Oligonucleotides were synthesized from a partial N-terminal sequence and used as probes in a polymerase chain reaction to isolate the gene from a genomic DNA digest. The gene was cloned in plasmid pBluescript II and sequenced. The resulting DNA-derived amino acid sequence was verified using a 41 and a 28 amino acid fragment. Comparison to known adenylate-kinase sequences revealed a glycin-rich P-loop. The number of identical positions in an **alignment** of 10 sequences was only 10; when **aligned** with eukaryotic sequences only, 30 identities were found. Secondary **structure** predictions exhibited notable similarity to determined helix/strand patterns of the pig or Escherichia coli enzyme. Homology modeling studies allowed the construction of **energy minimized** spatial **alignments** of large domains to known 3-D structures, suggesting that those were also resembled in the hyperthermophilic **protein**. Further studies on expression, mutagenesis and structural properties of the archael enzyme were discussed. (1 ref)

L29 ANSWER 38 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
DUPLICATE 26
ACCESSION NUMBER: 1992:162159 BIOSIS
DOCUMENT NUMBER: BA93:84484
TITLE: ENERGY-OPTIMIZED **STRUCTURE** OF ANTIFREEZE PROTEIN AND ITS BINDING MECHANISM.
AUTHOR(S): CHOU K-C
CORPORATE SOURCE: COMPUTATIONAL CHEM., UPJOHN RES. LAB., KALAMAZOO, MICH. 49001.
SOURCE: J MOL BIOL, (1992) 223 (2), 509-518.
CODEN: JMOBAK. ISSN: 0022-2836.
FILE SEGMENT: BA; OLD

LANGUAGE: English

AB A combination of Monte Carlo simulated annealing and **energy minimization** was utilized to determine the conformation of the antifreeze **protein** from the fish winter flounder. It was found from the energy-optimized **structure** that the hydroxyl groups of its four threonine residues, i.e. Thr2, Thr13, Thr35, are **aligned** on almost the same line parallel to the helix axis and separated successively by 16.1, 16.0 and 16.2 .ANG., respectively, very close to the 16.6 .ANG. repeat spacing along [0112] in ice. Based on such a space match, a zipper-like model is proposed to elucidate the binding mechanism of the antifreeze **protein** to ice crystals. According to the current model, the antifreeze **protein** may bind to an ice nucleation **structure** in a zipper-like fashion through hydrogen bonding of the hydroxyl groups of these four Thr residues to the oxygen atoms along the [0112] direction in ice lattice, subsequently stopping or retarding the growth of ice pyramidal planes so as to depress the freeze point. The calculated results and the binding mechanism thus derived accord with recent experimental observations. The mechanistic implications derived from such a special antifreeze molecule might be generally applied to elucidate the **structure**-function relationship of other antifreeze proteins with the following two common features: (1) recurrence of a Thr residue (or any other polar amino acid residue whose side-chain can form a hydrogen bond with water) in an 11-amino-acid period along the sequence concerned; and (2) a high percentage of Ala residue component therein. Further experiments are suggested to test the ice binding model.

L29 ANSWER 39 OF 44 MEDLINE on STN DUPLICATE 27

ACCESSION NUMBER: 93066156 MEDLINE
DOCUMENT NUMBER: 93066156 PubMed ID: 1438177
TITLE: Comparative molecular modeling and crystallization of P-30 protein: a novel antitumor protein of Rana pipiens oocytes and early embryos.
AUTHOR: Mosimann S C; Johns K L; Ardelt W; Mikulski S M; Shogen K; James M N
CORPORATE SOURCE: Medical Research Council of Canada, Department of Biochemistry, University of Alberta, Edmonton.
SOURCE: PROTEINS, (1992 Nov) 14 (3) 392-400.
Journal code: 8700181. ISSN: 0887-3585.
PUB. COUNTRY: United States
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 199212
ENTRY DATE: Entered STN: 19930122
Last Updated on STN: 19930122
Entered Medline: 19921223

AB The P-30 protein (Onconase) of Rana pipiens oocytes and early embryos is homologous to members of the pancreatic ribonuclease superfamily and exhibits an antitumor activity in vitro and in vivo. It appears that the ribonucleolytic activity of P-30 protein may be required for its antitumor effects. A comparative molecular model of P-30 protein has been constructed based upon the known three-dimensional **structure** of bovine pancreatic RNase A in order to provide structural information. Functionally, these enzymes hydrolyze oligoribonucleotides to pyrimidine-3'-phosphate monoesters and 5'-OH ribonucleotides. In the modeling procedure, automated sequence **alignments** were revised based upon the inspection of the RNase A **structure** before the amino acids of the P-30 **protein** were assigned the coordinates of the RNase A template. The inevitable intermolecular steric clashes that result were relieved on an interactive graphics device through the adjustment of side chain torsion angles. This process was followed by **energy minimization** of the model, which served to optimize stereochemical geometry and to relieve any remaining unacceptably close contacts. The resulting model retains the essential features of

RNase A as sequence insertions and deletions are almost exclusively found in exposed surface loops. The all atom superposition of active site residues of the P-30 protein model and an identically minimized RNase A **structure** has a root mean square deviation of 0.52 A. Though tentative, the model is consistent with a pyrimidine specificity.(ABSTRACT TRUNCATED AT 250 WORDS)

L29 ANSWER 40 OF 44 EMBASE COPYRIGHT 2003 ELSEVIER SCI. B.V. on STN

ACCESSION NUMBER: 92111169 EMBASE
DOCUMENT NUMBER: 1992111169
TITLE: Molecular conformation of ubiquitinated structures and the implications for regulatory function.
AUTHOR: Chun P.W.; Jou W.S.
CORPORATE SOURCE: Department of Biochemistry and Molecular Biology, College of Medicine, Box J-245, J.H.M.H.C., University of Florida, Gainesville, FL 32610, United States
SOURCE: Journal of Molecular Graphics, (1992) 10/1 (7-11+18). ISSN: 0263-7855 CODEN: JMGRDV
COUNTRY: United Kingdom
DOCUMENT TYPE: Journal; Article
FILE SEGMENT: 029 Clinical Biochemistry
LANGUAGE: English
SUMMARY LANGUAGE: English

AB The molecular conformation of ubiquitinated structures and the validity of the N-end rule were examined by simulating the molecular mechanics to ascertain the global **energy-minimized structure**. We examined the chemical linkage involved in attaching the ubiquitin carboxyl terminus to the N-terminus of three different x-hexapeptides, where x is the amino group of the acceptor **peptide** -either valine, arginine or glutamic acid-(x-K linkage) and to the .epsilon.-amino group of lysine of the acceptor **hexapeptide** x-glu1-his2-lys3-gly4-lys5-val6 (K-K linkage) through the formation of an **isopeptide** bond. Changes in conformation and molecular stability of the multi-ubiquitinated structures were determined by **energy-minimization** procedures using the SYBYL program developed by Tripos Associates. In the x-K linkage, the ubiquitin molecule is stretched in the .beta.-pleated sheets and .beta.-turns while the .alpha.-helices expand, as the molecule continues to unfold linearly. In the K-K linkage, the ubiquitin molecules have turned into a u-shaped, semicircular **alignment**, contracting into a compact, folded **structure**.

L29 ANSWER 41 OF 44 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN DUPLICATE 28

ACCESSION NUMBER: 1992:123713 BIOSIS
DOCUMENT NUMBER: BA93:69513
TITLE: THREE-DIMENSIONAL MODEL FOR STELLACYANIN A BLUE COPPER-PROTEIN.
AUTHOR(S): FIELDS B A; GUSS J M; FREEMAN H C
CORPORATE SOURCE: DEP. INORGANIC CHEM., UNIVERSITY SYDNEY, N.S.W. 2006, AUSTRALIA.
SOURCE: J MOL BIOL, (1991) 222 (4), 1053-1066. CODEN: JMOBAK. ISSN: 0022-2836.
FILE SEGMENT: BA; OLD
LANGUAGE: English

AB A three-dimensional model of the "blue" copper-glycoprotein stellacyanin from Rhus vernicifera has been derived by computer graphics, **energy minimization** and molecular dynamics techniques. The initial atomic co-ordinates were obtained by making substitutions and insertions in the known **structure** of another blue copper-**protein**, cucumber basic **protein** (CBP), which is 46% homologous with stellacyanin and has similar spectroscopic properties. An important difference between CBP and stellacyanin is that the latter lacks methionine, a residue that forms an exceptionally long bond to the copper atom in all blue copper-proteins of known **structure**. In the

aligned amino acid sequences, stellacyanin has glutamine 97 at the position that corresponds to the copper-binding methionine 89 in CBP. The hypothesis that the copper atom in stellacyanin is co-ordinated by the side-chain functional groups of histidine 46, cysteine 87, histidine 92 and glutamine 97 leads to a model that enables the spectroscopic properties, redox potential and electron-transfer kinetics of the **protein** to be rationalized. The present model for stellacyanin is more plausible than an antecedent model derived from the **structure** of plastocyanin. This demonstrates that the output from molecular modeling calculations is strongly dependent on the input, and that sequence homology with the target molecule is an important criterion for the selection of a starting model.

L29 ANSWER 42 OF 44 MEDLINE on STN DUPLICATE 29

ACCESSION NUMBER:	91129294 MEDLINE
DOCUMENT NUMBER:	91129294 PubMed ID: 2126456
TITLE:	Protein **structure** prediction.
AUTHOR:	Garnier J
CORPORATE SOURCE:	Unite d'Ingenierie des Proteines Biotechnologies, INRA, Jouy-en-Josas, France.
SOURCE:	BIOCHIMIE, (1990 Aug) 72 (8) 513-24. Ref: 71 Journal code: 1264604. ISSN: 0300-9084.
PUB. COUNTRY:	France
DOCUMENT TYPE:	Journal; Article; (JOURNAL ARTICLE) General Review; (REVIEW) (REVIEW, TUTORIAL)
LANGUAGE:	English
FILE SEGMENT:	Priority Journals
ENTRY MONTH:	199103
ENTRY DATE:	Entered STN: 19910405 Last Updated on STN: 19910405 Entered Medline: 19910320

AB Current methods developed for predicting protein **structure** are reviewed. The most widely used algorithms of Chou and Fasman and Garnier et al for predicting secondary **structure** are compared to the most recent ones including sequence similarity methods, neural network, pattern recognition or joint prediction methods. The best of these methods correctly predict 63-65% of the residues in the database with cross-validation for 3 conformations, helix, beta strand and coli with a standard deviation of 6-8% per protein. However, when a homologous protein is already in the database, the accuracy of prediction by the similarity peptide method of Levin and Garnier reaches about 90%. Some conclusions can be drawn on the mechanism of protein folding. As all the prediction methods only use the local sequence for prediction (+/- 8 residues maximum) one can infer that 65% of the conformation of a residue is dictated on average by the local sequence, the rest is brought by the folding. The best predicted proteins or peptide segments are those for which the folding has less effect on the conformation. Presently, prediction of tertiary **structure** is only of practical use when the **structure** of a homologous protein is already known. Amino acid **alignment** to define residues of equivalent spatial position is critical for modelling of the **protein**. We showed for serine proteases that secondary **structure** prediction can help to define a better **alignment**. Non-homologous segments of the polypeptide chain, such as loops, libraries of known loops and/or **energy minimization** with various force fields, are used without yet giving satisfactory solutions. An example of modelling by homology, aided by secondary **structure** prediction on 2 regulatory proteins, Fnr and FixK is presented.

L29 ANSWER 43 OF 44 EMBASE COPYRIGHT 2003 ELSEVIER SCI. B.V. on STN

ACCESSION NUMBER:	90109125 EMBASE
DOCUMENT NUMBER:	1990109125
TITLE:	Understanding structural relationships in proteins of

unsolved three-dimensional **structure**.

AUTHOR: Burbaum J.J.; Starzyk R.M.; Schimmel P.

CORPORATE SOURCE: Dept. of Biology, M.I.T. 16-437, 77 Massachusetts Ave.,Cambridge, MA 02139, United States

SOURCE: Proteins: Structure, Function and Genetics, (1990) 7/2 (99-111).
ISSN: 0887-3585 CODEN: PSFGEY

COUNTRY: United States

DOCUMENT TYPE: Journal; General Review

FILE SEGMENT: 029 Clinical Biochemistry

LANGUAGE: English

SUMMARY LANGUAGE: English

AB The locations of functionally important sequences and general structural motifs have been assigned to Ile-tRNA synthetase. However, a function has not been established for some segments of the **protein** (e.g., CP1). The method of structural modeling described here cannot establish the details of a 3 .ANG. crystal **structure**, and, in contrast to a crystal **structure**, the precision of the model varies according to the extent of a sequence similarity or the functional importance of a region. In Ile-tRNA synthetase, the signature sequence and the flanking regions are likely to be similar in **structure** to the proteins on which the model is based. For other regions, it may be possible to build a three-dimensional model by connecting well defined regions and refining the positions of the connecting elements by **energy minimization**. Structural modelling of this kind must be done cautiously, because the order and orientation of the elements of a structural motif can change in subtle ways. In the case of Tyr-tRNA synthetase, the .beta.-strand nearest the N-terminus is the outermost strand of the nucleotide binding fold; in Met-tRNA synthetase, the same strand is innermost. Furthermore, the orientation of this strand may be antiparallel (Tyr-tRNA synthetase) or parallel (Met-tRNA synthetase). Because multiple structures that differ in their orientations of structural elements are possible, the structural analogies between proteins should not be naively extrapolated without independent experimental support. As described above, some regions of proteins tolerate internal deletions and insertions. This provides further experimental support for the practice of allowing for gaps in computer-generated sequence **alignments**. Nevertheless, because some regions are more tolerant of insertions and deletions than others, the structural and functional significance of a region of broken **alignment** must be assessed carefully. All gaps in sequence **alignments** cannot be treated equally, and each must be evaluated within its own context. In the synthetases of known **structure**, structural analogy can be used to identify important functional elements. For example, the amino acid binding site of Met-tRNA synthetase might be formed, at least in part, by a **peptide** that encompasses Ala50; this amino acid **aligns** with Gly94 of the Ile-tRNA synthetase. This is an example in which results on a **protein** of unknown **structure** (Ile-tRNA synthetases) can lead to identification of a potential substrate binding site in a **protein** of known **structure** (Met-tRNA synthetase).

L29 ANSWER 44 OF 44 BIOSIS
DUPLICATE 30

ACCESSION NUMBER: 1985:233326 BIOSIS

DOCUMENT NUMBER: BA79:13322

TITLE: A REFINED MODEL FOR THE VARIABLE DOMAINS FV OF THE J-539 BETA-1 6-D GALACTAN-BINDING IMMUNOGLOBULIN.

AUTHOR(S): MAINHART C R; POTTER M; FELDMANN R J

CORPORATE SOURCE: NATL. CANCER INST., BETHESDA, MD 20205.

SOURCE: MOL IMMUNOL, (1984) 21 (6), 469-478.
CODEN: MOIMD5. ISSN: 0161-5890.

FILE SEGMENT: BA; OLD

LANGUAGE: English

AB A refined protocol for building a hypothetical model of the [mammalian] J539 Fv is described. Computer programs for positioning amino acid side chains and **structure energy minimization** were employed. Computer modeling was accomplished on an Evans and Sutherland picture system which permitted **structure** visualization in 3 dimensions. **Peptide** backbone breaksites were rejoined by monitoring for correct distances and torsion angles. A physical model was then constructed and used as a basis for further refinements such as **aligning** conformations around remodeled sites, adjusting proline substitutions and optimizing H-bond-forming potentials. This **structure** (J539-ADO) was **energy minimized**; the final coordinates were obtained from the energy-refined model. The resulting hypothetical J539 **structure** can be compared to the **structure** of J539 now being determined by X-ray crystallography. The procedures described can be used for other Fv fragments.

```
=> d his

     (FILE 'HOME' ENTERED AT 09:50:49 ON 15 AUG 2003)

     FILE 'MEDLINE, BIOSIS, BIOTECHDS, CAPLUS, EMBASE' ENTERED AT 09:51:02 ON
     15 AUG 2003
                E BLANKENBECLER RICHARD/IN
L1            5 S E3
                E OHLSSON M
                E OHLSSON MATTIAS
                E OHLSSON M/IN
L2        21855 S PROTEIN (S) ALIGN?
L3        24619 S (PROTEIN OR ?PEPTIDE) (S) ALIGN?
L4        10381 S STRUCTURE (A) L3
L5         1253 S STRUCTURE (S) ALGIN?
L6            4 S L4 AND L5
L7        18438 S STRUCTURE (S) ALIGN?
L8         6639 S L4 AND L7
L9          481 S ATOMIC DISTANCE
L10        1596 S STOM? AND DISTANCE?
L11        1740 S ATOM? DISTANCE?
L12           4 S L8 AND L11
L13       20394 S MEAN FIELD
L14          14 S L4 AND L13
L15           8 DUP REM L14 (6 DUPLICATES REMOVED)

     FILE 'STNGUIDE' ENTERED AT 10:25:18 ON 15 AUG 2003
L16           0 S (BINARY OR POTTS) (S) ASSIGN
L17           0 S BINARY (S) ASSIGN?
L18           0 S BINARY ASSIGN?

     FILE 'MEDLINE, BIOSIS, BIOTECHDS, CAPLUS, EMBASE' ENTERED AT 10:32:21 ON
     15 AUG 2003
L19         689 S (BINARY OR POTTS) (S) ASSIGN?
L20       18438 S STRUCTURE (S) ALIGN?
L21           5 S L19 AND L20
L22           3 DUP REM L21 (2 DUPLICATES REMOVED)
L23          24 S (BINARY OR POTTS) (A) ASSIGN?
L24          12 DUP REM L23 (12 DUPLICATES REMOVED)
L25           2 S POTTS (A) ASSIGN?
L26        9254 S ENERGY (A) MINIMIZ?
L27         109 S L4 AND L26
L28          83 S L7 AND L27
L29          44 DUP REM L28 (39 DUPLICATES REMOVED)

=> s energy (a) cost
```

```
L30         6687 ENERGY (A) COST

=> s l30 and l29
L31            0 L30 AND L29

=> s structure (a) align?
L32         1327 STRUCTURE (A) ALIGN?

=> s l30 and l32
L33            0 L30 AND L32
```